Computational Modelling of Glycosidase Mechanisms: Structural and Mechanistic aspects

Soliman, Mahmoud

Award date:
2009

Awarding institution:
University of Bath

Computational Modelling of Glycosidase Mechanisms:

Structural and Mechanistic aspects

Mahmoud Elsayed Soliman Soliman

A thesis submitted for the degree of Doctor of Philosophy
University of Bath
Department of Chemistry
September/2009

Computational Modelling of Glycosidase Enzymes

CONTENTS

Chapter 5

Chapter 6

Chapter 7

Chapter 8

ABSTRACT

Computer simulation provides an excellent way to study enzyme reactions both to understand the fundamentals of catalysis and to aid the design of transition-state analogues as mechanism-based inhibitors. In order to describe the electronic processes involved in chemical bond making and breaking, while at the same time treating realistically the complex environments of proteins in solution, we employ hybrid QM/MM methods: the catalytic residues and the substrate are described quantum mechanically while their surroundings are represented by molecular mechanics. Moreover, molecular dynamics offers a means to sample the very large numbers of thermally accessible configurations representative of these large molecular ensembles at ambient temperatures, with appropriate simulations enabling evaluations of free-energy profiles and free-energy surfaces.

The family GH-11 endo-1,4-β-xylanase from *B. circulans* hydrolyses xylan with net retention of anomeric configuration by a double-displacement mechanism involving Glu78 as a nucleophile and Glu172 as a general acid. The glycone adopts a $^{2,5}B$ boat conformation in the active site. Molecular dynamics simulations have been performed for non-covalent complexes of phenyl β-xylobioside with the retaining endo-β-1, 4-xylanase from *B. circulans* (BCX) and its Tyr69Phe mutant using a hybrid QM/MM methodology. A trajectory initiated for the wild-type enzyme-substrate complex with the proximal xylose ring bound at the −1 subsite (adjacent to the scissile glycosidic bond) in the $^{4}C_{1}$ chair conformation shows spontaneous transformation to the $^{2,5}B$ boat conformation, and potential of mean force calculations indicate that the boat is ~30 kJ mol^{-1} lower in free energy than the chair. Analogous simulations for the mutant lacking one oxygen atom confirm the key role of Tyr69 in stabilizing the boat in preference to the $^{4}C_{1}$ chair conformation, with a relative free energy difference of about 20 kJ mol^{-1}, by donating a hydrogen bond to the endocyclic oxygen of the proximal xylose ring. QM/MM MD simulations for phenyl β-xyloside in water, with and without a propionate/propionic acid pair to mimic the catalytic glutamate/glutamic acid pair of the enzyme, show the $^{4}C_{1}$ chair to be stable, although a hydrogen bond between the OH group at C2 of xylose and the propionate moiety seems to provide some stabilization for the $^{2,5}B$ conformation.

Two-dimensional potential-of-mean-force calculations based upon molecular dynamics with the AM1/OPLS method for wild-type BCX with a p-nitrophenyl xylobioside substrate in water clearly indicate a stepwise mechanism for glycosylation: the rate-determining step is nucleophilic substitution by Glu78 to form the covalently bonded enzyme-substrate intermediate without protonation of the leaving group by Glu172. The geometrical configuration of the transition state for the enzymic reaction is essentially the same as found for a gas-phase model involving only the substrate and a propionate/propionic acid pair to represent the catalytic glutamate/glutamic acid groups. In addition to stabilizing the $^{2,5}B$ boat conformation of the proximal xylose in the non-covalent reactant complex of the substrate with BCX, Tyr69 lowers the AM1/OPLS free-energy barrier for glycosylation by 42 kJ mol^{-1} relative to that calculated for the Y69F mutant, which lacks the oxygen atom. In the oxacarbenium ion-like transition state the phenolic oxygen of Tyr69 approaches closer to the endocyclic oxygen of the sugar ring but donates its hydrogen bond not to the ring oxygen but rather to the nucleophilic oxygen of Glu78. Comparison of the average atomic charge distributions for the wild-type and mutant indicates that charge separation along the bond between the anomeric carbon and the ring oxygen is matched in the former by a complementary separation of charge along the Tyr69 O---H bond, corresponding to a pair of roughly antiparallel bond dipoles, which is not present in the latter.

The calculated free energy barrier for the nucleophilic attack step for Glu78/Asp mutant is higher by 40 kJ mol^{-1} relative to that calculated for the wild-type BCX. Thus the proper positioning of the nucleophilic residue is so crucial for the catalytic process.

QM/MM kinetic isotope effects on V_{max}/K_m (2 α-$^{2}H_3$, β-^{2}H, γ-$^{2}H_2$ and $^{18}O_{ring}$ and 1 α-^{13}C and $^{18}O_{lg}$) have been computed by means of semi-classical partition functions averaged over many thermally accessible configurations of the free substrate in water and the wild-type enzymatic transition state, generated by MD simulations with periodic boundary conditions.

ACKNOWLEDGEMENTS

I am truly grateful to everyone who has directly or indirectly supported and helped me on the occasion of the completion of the thesis.

I find immense pleasure in expressing my sincere regards and profound gratitude to my research supervisor Prof. Ian H. Williams, professor of theoretical chemistry, chemistry department, university of bath, for his guidance. I owe the greatest depth of gratitude for his invaluable help, strong motivation, encouragement and support. Apart from his guidance, he has given me the experience, built confidence and lifted my spirits, which cannot be described in words. All of that was very important for somebody who has no previous knowledge about computational chemistry like me.

I wish to thank Jan, Ian Williams' wife, for her hospitable attitude and for the nice English food she makes either during the research group invitations or family invitations.

I take immense pleasure to offer my special thanks to postdocs Dr. Giuseppe D. Ruggiero, J. Javier Ruiz Pernía and Ian R. Greig for their constant encouragement and suggestions that have been valuable throughout the research.

I am indebted to express my loving thanks to my wife, Shimaa, for supporting me and looking after my two noisy kids, Adham and Tameem, during all the last four years.

I would like to thank the whole scientific as well as non scientific staffs of chemistry division in university of bath and also the IT technician, Mark Russell, for supporting my thesis directly or indirectly.

I also thank The Egyptian Government for financial assistance.

Mahmoud

1. Subject of the thesis

1.1. Foreword

Chemistry has often been referred to as "The Central Science." Knowledge of the properties and behaviour of atoms and molecules is crucial to our understanding of reaction mechanisms, catalysis, biological systems and many other fields.

In silico chemistry has appeared in the last few years, in analogy with the *in vivo / in vitro* designations used for chemical/biochemical results and also other techniques like bioinformatics, screening methods and rational drug design. Starting from simple (almost trivial) pairwise interactions of atoms, we can simulate complex molecular systems and biomolecules on longer timescales and study features on the limit of state-of-the-art experiments.

Biosystems are large by nature. In the past, classical molecular mechanics (MM) took place to describe them. Nevertheless, a large number of characteristics cannot be covered by classical treatment. However, although quantum mechanics (QM) can give us a complete and accurate picture and can inspire improvements in molecular mechanics, the computational demand will be too great. Quantum mechanical/molecular mechanical (QM/MM) methods allow reactions in large systems to be studied. A small region including the reacting atoms is treated with a QM method, while most of the system is represented more simply by molecular mechanics.

1.2. The aim and contents of the thesis

The main core of this study is to provide further and deeper understanding of the mechanistic and structural features of endo-β-1,4-xylanase from *Bacillus circulans*. Wherever possible, theoretical results are compared with experimental data.

This thesis presents the main results of my research performed in Prof. Ian H. Williams' group during the period 2005–2009.

The content is divided into eight chapters, where the first constitutes a background to the field of research and puts the results in a larger context. *Chapter 2* is an introduction to glycosidase enzymes in general, while *Chapter 3* provides a detailed overview on the enzyme of interest, endo-β-1,4-xylanase from *Bacillus circulans*, in particular. *Chapter 4* outlines the objectives of this study. From *chapter 5* until *chapter 7*, each chapter is considered as a separate unit in terms of the problem I probe, computational approaches I apply, and results I obtain. *Chapter 8* addresses the ongoing work and future plans of our research.

2. Publications come out of this work

2.1. Published

1. Soliman, M. E. S., Ruggiero, G. D., Ruiz Pernía, J. J., Greig, I. R., and Williams, I. H. (2009), Computational mutagenesis reveals the role of active-site tyrosine in stabilising a boat conformation for the substrate: QM/MM molecular dynamics studies of wild-type and mutant xylanases. *Org. Biomol. Chem.*, **7**, 460-468.
2. Soliman, M. E. S., Ruiz Pernía, J. J., Greig, I. R., and Williams, I. H. (2009), Mechanism of glycoside hydrolysis: A comparative QM/MM molecular dynamics analysis for wild type and Y69F mutant retaining xylanases. *Org. Biomol. Chem.*, **xxxxxxxx**.

2.2. Manuscripts in preparation

1. Soliman, M. E. S., Ruiz Pernía, J. J., Greig, I. R., and Williams, I. H. (2009), Kinetic isotope effect studies for endo-β-1, 4-xylanase from B. *circulans* (BCX).
2. Soliman, M. E. S., Ruiz Pernía, J. J., Greig, I. R., and Williams, I. H. (2009), positioning of the catalytic residues among glycosidases; updated survey and case study.

3. Soliman, M. E. S., Ruiz Pernía, J. J., Greig, I. R., and Williams, I. H. (2009), the effect of variation of the pK_a of leaving group on the catalytic mechanism for BCX; a QM/MM/MD study.
4. Soliman, M. E. S., Ruiz Pernía, J. J., Greig, I. R., and Williams, I. H. (2009), QM/MM/MD insight on deglycosylation route for BCX.

Introduction to computational chemistry

In this chapter, I attempt to provide a broad overview of the methods that have been used during this research, along with a wider survey of other methods used by computational chemists.

I would like to point out the difference between the theoretical and computational chemist. Simply, the former develops and implements the tools while the latter applies standard techniques to solve relevant chemical problems. In other words, theoretical chemistry may be defined as the mathematical description of chemistry while computational chemistry is usually used when a mathematical method is sufficiently well developed that it can be automated for implementation on a computer.

In the first section I handle a general but likely particular overview of techniques and strategies in theoretical chemistry. In the later sections I provide a deeper explanation of the methods used in this thesis. Over the period of my research I have been guided by many good books, articles and papers, which you can find in the reference section.

1.1 A historic overview of the theoretical methods

The core of theoretical chemistry is theoretical physics, developed when physicists in the early 20th century adopted quantum mechanics to molecular systems. Understanding of chemical bonds, electromagnetic spectra, and the explanation of stability of some molecular structures were some of the greatest achievements. A lot of effort has been done since then to improve the initial work made by Slater, Mulliken, Pauling, London and many others. This is why historically quantum chemistry has almost taken the whole task in theoretical chemistry (*1*).

1.2 Theoretical methods used in this thesis:

Potential energy

Quantum mechanics can be an ideal and powerful tool for simulation of chemical reactions. In practice, very small systems can be solved by means of quantum theory. Generally, chemical potential energy is a form of potential energy related to the structural arrangement of atoms or molecules. This arrangement may be the result of chemical bonds within a molecule. Potential energy can be divided into electronic and nuclear potential energy. The potential energy that gives title to this section is the potential that molecular nuclei feel under the kinetic and potential energy influence of electrons. The Born-Oppenheimer approximation is a mathematical strategy, which deals with the separation of nucleus and electronic motion, is considered the conceptual foundation of the quantum methods that will be discussed in details in section 1.2.2.

1.2.1 Potential energy surface

From a computational point of view, many aspects of chemistry can be reduced to questions about potential energy surfaces (PES). A model surface of the energy as a function of the molecular geometry is shown in Figure 1.1 to illustrate some of the features. One can think of it as a hilly landscape, with valleys, mountain passes and peaks. Molecular structures correspond to the positions of the minima in the valleys on a potential energy surface. The energetics of reactions are easily calculated from the relative energies or altitudes of the minima for reactants and products. Reaction rates can be obtained from the height and profile of the mountain pass separating the valleys of the reactants and products.

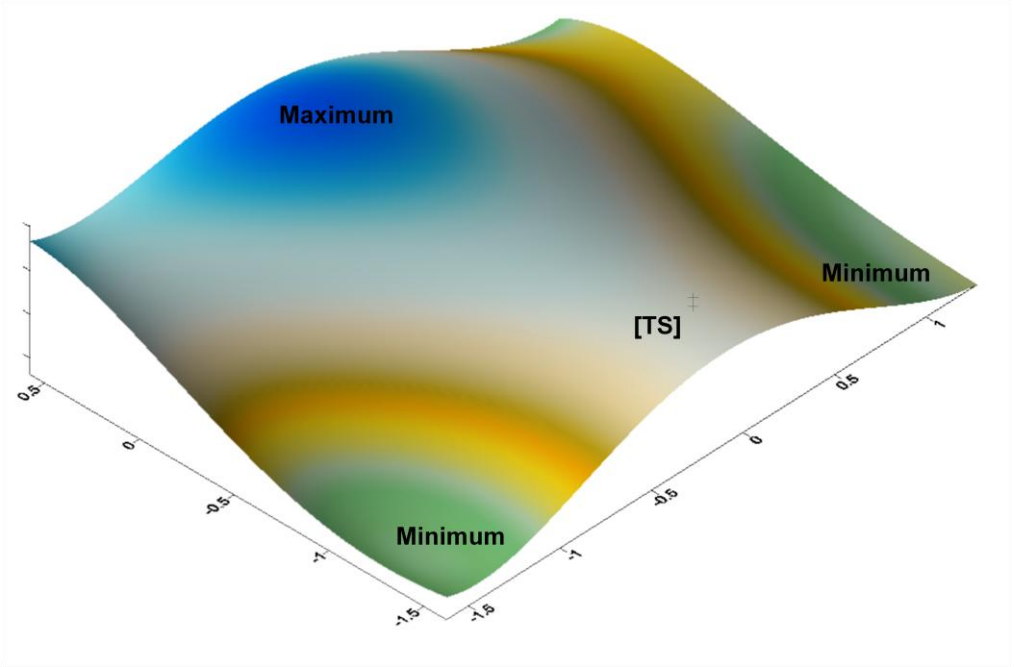


Fig 1.1. Example of an energy landscape produced by the systematic alteration of two degrees of freedom (two-dimensional PES).

A potential energy surface (PES) for a collection of *N* atoms can be considered as a multidimensional hypersurface that is dependent on the resulting 3*N*-6 internal degrees of freedom associated with that system. An energy change occurs when any of the degrees of freedom is altered and for most systems, it is not usually viable, or necessary to investigate the entire hypersurface. Subsequently, an energy landscape can be generated from the systematic alteration of two independent geometrical parameters. As the value of each parameter is changed, the energy is minimized in all the other degrees of freedom. In such case we will get what so called a two-dimensional PES.

On the surface, there exist a number of stationary points. Minima usually correspond to equilibrium geometries, such as reactants and products on a reaction profile. The route taken across the surface for any reaction is the lowest energy pathway possible, with reactants and products located at minima on either side of the surface. For a chemical step such as that shown in Figure 1, there is a maximum on the lowest energy pathway corresponding to a *first-order saddle point*. The structure at such a point is known as the *transition structure* (TS) and it determines the energy barrier that must be overcome when proceeding from reactants to products.

For any stationary point on the PES, the energy change with respect to all the variables (i.e. the force) is zero. For *i* variables therefore:

$$\left(\frac{\partial E}{\partial R_1},\frac{\partial E}{\partial R_2},\frac{\partial E}{\partial R_3},\frac{\partial E}{\partial R_4},.....,\frac{\partial E}{\partial R_i}\right)=0 \qquad (1.1)$$

This verifies the presence of a stationary point on the PES. To figure out the nature of such a point, the second derivatives of the energy which lead to an *i* x *i* matrix known as the *hessian* should be identified. *Hessian* matrix takes the following format:

$$\begin{pmatrix} \frac{\partial^2 E}{\partial R_1^2} & \frac{\partial^2 E}{\partial R_1 \partial R_2} & \frac{\partial^2 E}{\partial R_1 \partial R_3} & \frac{\partial^2 E}{\partial R_1 \partial R_4} & ..... & \frac{\partial^2 E}{\partial R_1 \partial R_i} \\ \frac{\partial^2 E}{\partial R_2 \partial R_1} & \frac{\partial^2 E}{\partial R_2^2} & \frac{\partial^2 E}{\partial R_2 \partial R_3} & \frac{\partial^2 E}{\partial R_2 \partial R_4} & ..... & \frac{\partial^2 E}{\partial R_2 \partial R_i} \\ \frac{\partial^2 E}{\partial R_3 \partial R_1} & \frac{\partial^2 E}{\partial R_3 \partial R_2} & \frac{\partial^2 E}{\partial R_3^2} & \frac{\partial^2 E}{\partial R_3 \partial R_4} & ..... & \frac{\partial^2 E}{\partial R_3 \partial R_i} \\ ..... & & & & & \\ ..... & & & & & \\ \frac{\partial^2 E}{\partial R_i R_1} & \frac{\partial^2 E}{\partial R_i \partial R_2} & \frac{\partial^2 E}{\partial R_i \partial R_3} & \frac{\partial^2 E}{\partial R_i \partial R_4} & ..... & \frac{\partial^2 E}{\partial R_i^2} \end{pmatrix} \qquad (1.2)$$

Under the harmonic approximation (i.e. a symmetrical potential energy well), the force F_i, on a given geometrical variable, R_i, is given by:

$$F_i = \kappa_i R_i \qquad (1.3)$$

Where κ_i is the force constant. If R_i represents a bond, then the force constant is a measure of the bond strength. A high value for the force constant represents a stiff bond (with a narrow potential) while smaller force constant results in a weaker and broader potential (Figure 1.2).

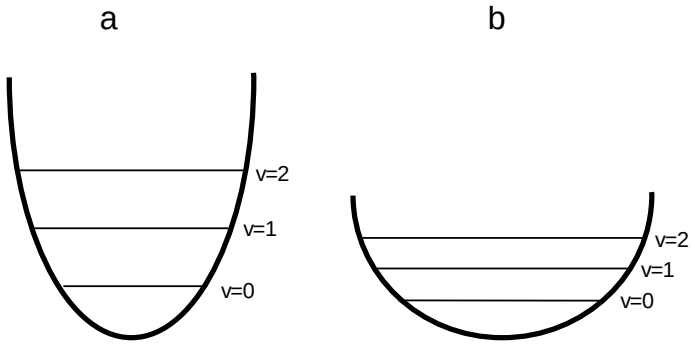


Fig 1.2. Schematic diagram showing a). a tight potential (large force constant) and b). a loose potential (small force constant). Both situations assume a harmonic potential; this is appropriate for the base of the potential energy well.

The second derivative of energy with respect to the geometrical parameters leads to the force constants. For a diatomic molecule, the vibrational frequency, ν, associated with the bond stretch is given by eq. 1.4, where μ is the reduced mass of the molecule.

$$\nu = \frac{1}{2\pi}\sqrt{\frac{\kappa}{\mu}} \tag{1.4}$$

For an *N* atom system, the hessian matrix in Cartesian co-ordinates has $3N \times 3N$ elements. The vibrational frequencies are obtained from the eigenvalues of the mass-weighted hessian and the vibrational modes are the corresponding eigenvectors. For a non-linear molecule (located at a stationary point), there are 6 zero eigenvalues corresponding to the overall translational and rotational motion of the molecule in three-dimensional space.

For minima stationary points, the energy is a minimum in all directions; the resulting eigenvalues of the hessian are all positive and the vibrational frequencies have real values. The first-order saddle point, however, is a maximum in one direction and a minimum in all the others. At this point, therefore, one of the eigenvalues in the hessian is negative and the associated vibrational frequency is imaginary (since the square root of the negative eigenvalues is used to evaluate the frequency).

Big molecules such as enzymes and other condensed phase systems are too big to be computationally affordable with quantum methods, even to compute only

their potential energy surface. To overcome these limitations what so-called Quantum Mechanics / Molecular Mechanics (QM/MM) methods have been one of the most successful applications in recent years.

In order to explain the QM/MM methods used in this thesis, it is important to overview the currently available methods in Quantum Mechanics and Molecular Mechanics, and their advantages and its drawbacks. In this way, we can justify the appearance and the success of hybrid methods in general and of QM/MM in particular.

1.2.2. Quantum Mechanics

The word “quantum” came from the Latin word which means "what quantity". In quantum mechanics, it refers to a discrete unit that quantum theory assigns to certain physical quantities, such as the energy of an atom at rest. The discovery that waves have discrete energy packets (called quanta) that behave in a manner similar to particles led to the branch of physics that deals with atomic and subatomic systems which we today call quantum mechanics. It is the underlying mathematical framework of many fields of physics and chemistry, including condensed matter physics, solid-state physics, atomic physics, molecular physics, computational chemistry, quantum chemistry, particle physics, and nuclear physics *(2-4)*.

1.2.2.1. Basic equations for a molecular system

An electron has wave-like properties whose nature is determined by its kinetic energy and interaction with other particles in the system; such properties are described by Schrödinger equation which describes the evolution of the quantum system:

$$-\frac{h}{i}\frac{\delta|\psi\rangle}{\delta t}=H|\psi\rangle \qquad (1.5)$$

Where $|\psi\rangle$ is the vector of the space that contains all the information of the system and H is the Hamiltonian operator (the sum of kinetic and potential energy). Under the representation of positions the wavefunction and for a molecular system that contains $\boldsymbol{N}$ electrons as well as $\boldsymbol{M}$ nuclei with charge $\boldsymbol{Z}$ this Hamiltonian has the following form:

$$H = T_M + T_N + V_{NM} + V_{NN} + V_{MM} \tag{1.6}$$

Or in a more detailed description

$$H = -\frac{1}{2}\sum_{K}^{M}\frac{\nabla_K}{m_K} - \frac{1}{2}\sum_{i}^{N}\nabla^2{}_i - \sum_{i}^{N}\sum_{K}^{M}\frac{Z_K}{r_{iK}} + \sum_{i}^{N}\sum_{i>j}^{N}\frac{1}{r_{ij}} + \sum_{K}^{M}\sum_{K>L}^{M}\frac{Z_K Z_L}{R_{KL}} \tag{1.7}$$

Where,

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{1.8}$$

The first and second terms in equation 1.6 are the kinetic energy of the nuclei and electrons, respectively, while the third, fourth and fifth terms represent electron-nuclear, electron-electron and nuclear-nuclear interactions, respectively.

Under the consideration that $|\psi\rangle$ is a stationary state we get the time independent Schrödinger equation

$$H|\psi\rangle = E|\psi\rangle \tag{1.9}$$

The value E is the energy eigenvalue of the Hamiltonian operator, a scalar that offers the spectrum of the operator.

The equation 1.9 cannot be solved exactly for a molecular system. Consequently we will need to solve the equation 1.9 by different stages. This two-stage solution is provided by Born-Oppenheimer approximation.

1.2.2.2. Born-Oppenheimer approximation:

Born-Oppenheimer approximation shows how the electronic motion can be approximately separated from the nuclear motion.

Let us define a molecule by a geometrical structure determined by the nuclear positions. If the nuclei have fixed positions, the nuclear kinetic term vanishes ($T_M = 0$) and the nuclear repulsion term V_{MM} becomes a constant. The Hamiltonian

expression of equation 1.6 has a shortened form that we label as the electronic Hamiltonian

$$H^{elec} = T_N + V_{NM} + V_{NN} \tag{1.10}$$

The solutions of the electronic Hamiltonian are the electronic wave functions that will be solved for every nuclear configuration R_K

$$H^{elec} \left| \Psi_{R_K}^{elec} \right\rangle = E_{R_K}^{elec} \left| \Psi_{R_K}^{elec} \right\rangle \tag{1.11}$$

The term V_{MM} is not usually included in the electronic Hamiltonian since it is a constant at a given nuclear configuration. However we can define the potential energy by adding the term V_{MM} to the electronic energy

$$U_{R_K} = E_{R_K}^{elec} + V_{MM} \tag{1.12}$$

The solutions $\{ \left| \Psi_{R_K}^{elec} \right\rangle \}$ are a complete set of functions of the *n*-electron space. So, the total wave function of the system should belong to a full space created from the tensorial product between the nuclear space and electronic space:

$$E = E^{elec} + \sum_{K}^{M} \sum_{K.L}^{M} \frac{Z_K Z_L}{R_{KL}} \tag{1.13}$$

In conclusion, the nuclei move on a potential energy surface (PES) ***U***, where ***U*** comes from the solution of the electronic Schrödinger equation 1.11, and ***E*** is the total energy of the system. The PES is a concept that we will be using throughout this thesis.

1.2.2.3 Electronic problem: Hartree-Fock

To understand the Hartree-Fock approximation we should first shed some light on some electronic aspects such as *electron-spin* and *antisymmetry*.

The *one-electron* Hamiltonian h for electron i in the molecular system is given by its kinetic energy and its potential energy of interaction with all nuclei.

$$h_i = -\frac{1}{2} \nabla_i^2 - \sum_{\alpha=1}^{M} \frac{Z_\alpha}{r_{i\alpha}} \tag{1.14}$$

Each electron is contained within one spin-molecular orbital, χ_i (where there are two spin molecular orbitals to a spatial molecular orbital). The energy of each individual electron is given by:

$$E = H_{ii}^{core} = \int \chi_i(h_i)\chi_i d\tau \tag{1.15}$$

H_{ii}^{core} represents the energy of the electron in a bare field of nuclei. In addition to these one-electron integrals, *two-electron integrals* account for the repulsion energy between electrons. Thus, the repulsion between two electrons with like-spin, electron 1 in spin orbital χ_i and electron 2 in spin orbital χ_j, is:

$$E = J_{ij} = \iint \chi_i(1)\chi_j(2)(\frac{1}{r_{12}})\chi_i(1)\chi_j(2)d\tau_1 d\tau_2 \tag{1.16}$$

However, there exists a small favourable energy term: electrons with the same spin are not permitted to occupy the same position in space, as stated by the *Pauli exclusion principle*. The subsequent reduced repulsion between electrons with the same spin-orientation manifests itself a favourable energy contribution and this is known as an *exchange interaction*:

$$E = K_{ij} = \iint \chi_i(1)\chi_j(2)(\frac{1}{r_{12}})\chi_i(1)\chi_j(2)d\tau_1 d\tau_2 \tag{1.17}$$

For a closed shell system, two electrons in one spatial orbital interact in four different ways with two electrons in another spatial orbital (Figure 1.3).

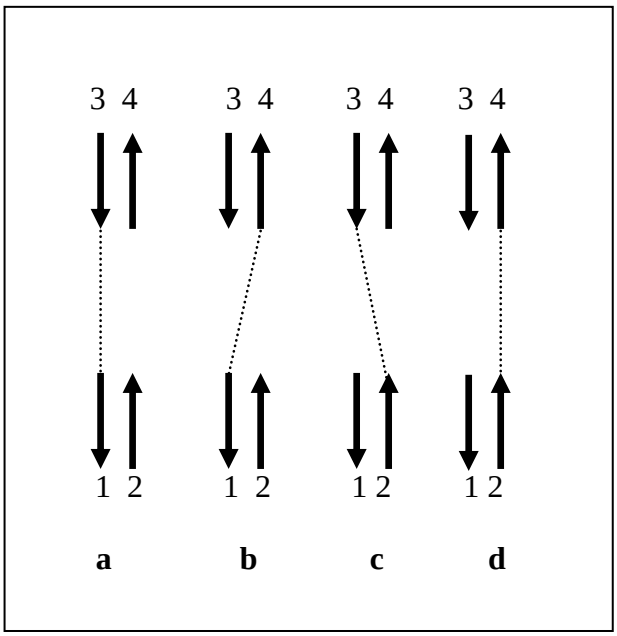


Fig 1.3. Various ways by which two electrons in one spatial orbital can interact with the electrons in another

The total repulsion energy between the two pairs of electrons in figure 1.3 is thus $4J_{ij}$; this is an over-estimate of the coulombic interactions because two of the situations involve unpaired spins (figure 1.3a, figure 1.3d). This necessitates inclusion of the exchange energy ($4K_{ij}$). Finally, electron-electron interactions, J_{ii}, within the same spatial orbital must also be taken into account. Thus for a closed shell system with *N*/2 filled spatial molecular orbitals, the total electronic energy of the system is:

$$E = 2\sum_{i=1}^{N/2} H_{ii}^{core} + \sum_{i=1}^{N/2}\sum_{j=i+1}^{N/2}(4J_{ij} - 2K_{ij}) + \sum_{i=1}^{N/2} J_{ii} \qquad (1.18)$$

Because $J_{ii} = K_{ii}$, this equation can be simplified by "double counting" all the electron-electron interactions found within the system:

$$E = 2\sum_{i=1}^{N/2} H_{ii}^{core} + \sum_{i=1}^{N/2}\sum_{j=i}^{N/2}(2J_{ij} - K_{ij}) \qquad (1.19)$$

This is the Hartree-Fock (HF) electronic energy for a closed shell system with *N* electrons occupying *N*/2 molecular orbitals. It should be noted that the overall HF energy would also account for the "constant" nuclear-nuclear repulsions.

Since the electrons are fermions they must be antisymmetric when interchanged. To antisymmetrize the Hartree product the total wavefunction is written in the form of a single determinant (a so-called Slater determinant) this means that it is antisymmetric upon interchange of electron coordinates.

$$\left|\Psi^{Slater}\right\rangle = \frac{1}{\sqrt{N!}}\begin{vmatrix} \chi_1(1)\,\chi_2(1)\ldots\ldots\,\chi_n(1) \\ \chi_1(2)\,\chi_2(2)\ldots\ldots\,\chi_n(2) \\ \\ \chi_1(N)\,\chi_2(N)\ldots\ldots\chi_n(N) \end{vmatrix} \tag{1.20}$$

The set of molecular orbitals leading to the lowest energy are obtained by a process referred to as a "self-consistent-field" or SCF procedure. The archetypal SCF procedure is the Hartree-Fock procedure. All SCF procedures lead to an equation of the form:

$$f(i)\chi(x_i) = \varepsilon\chi(x_i) \tag{1.21}$$

Here, $f(i)$ is the Fock operator which can be written as:

$$f(i) = -\frac{1}{2}\nabla_i^2 + \upsilon^{eff}(i) \tag{1.22}$$

x_i are spin and spatial coordinates of the electron i and υ^{eff} is the effective potential "seen" by the electron i, which depends on the spin orbitals of the other electrons.

1.2.2.4. LCAO Approximation

The Hartree-Fock approximation leads to a set of coupled differential equations (the Hartree-Fock equations), each involving the coordinates of a single electron. While they may be solved numerically, it is advantageous to introduce an additional approximation in order to transform the Hartree-Fock equation into a set of algebraic equations.

It is reasonable to expect that the one-electron solutions for many-electron molecules will closely resemble the (one-electron) solution for the hydrogen atom. After all, the molecules are made up of atoms, so why shouldn't molecular solutions made up of atomic solutions? In practice the spatial molecular orbitals,

ψ_i, are formulated as a *linear combination of atomic orbitals* (LCAO). Thus T atomic orbitals ϕ contribute to forming T molecular orbitals:

$$\psi_i = \sum_{j=1}^{T} c_j \phi_j \tag{1.23}$$

The constant, c_j determines the magnitude of contribution that the atomic orbitals ϕ_j makes to the eventual molecular orbital ψ_i. The atomic orbitals may also be represented by *basis functions*, and a complete description of the molecular orbital is obtained from an infinite number of basis functions (where $T = \infty$). Therefore, the more basis functions used to generate the molecular orbital, the more complete description of the system.

1.2.2.5. Introduction of a basis: Roothaan-Hall equations

The Hartree-Fock and LCAO approximations, taken together and applied to the Schrödinger equation, lead to the Roothaan-Hall equation:

$$Fc = \varepsilon Sc \tag{1.24}$$

Here ε are orbital energies, S is the overlap matrix (a measure of the extent to which basis functions "see each other"), and F is the Fock matrix, which is analogous to the Hamiltonian in the Schrödinger equation.

The basis sets used within this thesis consist of a set of gaussian functions (gf) of the form, exp ($-\alpha r^2$). The α term represents how broad the gf is, with smaller values giving a more diffuse function. Each basis function comprises of a linear combination of Gaussian functions: the use of individual gaussian functions provides a poor description of the electron density at the nucleus and at longer electron-nuclear distances. Thus:

$$\phi_i = \sum_{j=1}^{G} d_j \gamma_j \tag{1.25}$$

Where, γ_j are the gaussian functions used in generating atomic orbitals, ϕ_i, and d_i is the corresponding coefficient for each gf. To generate self-consistent molecular orbitals, the values of d_i and α are altered throughout the calculation for every gf in the basis set. The labeling of such a basis set is as follows:

W-XYZ (Dif1, Dif2) G (Po11, Po12)

W represents the number of gaussian functions describing the "inner" electrons (i.e. the 1s electrons for second row elements); for all basis sets used herein, W=6. The X, Y and Z basis function describe the valence shell electrons, in which the X function is contracted while Y and Z are more diffuse. This allows increased orbital flexibility when minimizing the energy of the system compared to only having one basis function alone. The *split-valence double-zeta* basis set describes the valence shell orbitals using two functions (namely, X and Y).

Today, there are hundreds of basis sets composed of gaussian functions. The smallest of these are called *minimal basis sets*, and they are typically composed of the minimum number of basis functions required to represent all of the electrons on each atom. The largest of these can contain literally dozens to hundreds of basic functions on each atom. A minimal basis set is one in which, on each atom in the molecule, a single basis function is used for each orbital in a Hartree-Fock calculation on the free atom. However, for atoms such as lithium, basis functions of p type are added to the basis functions corresponding to the 1s and 2s orbitals of the free atom. For example, each atom in the first row of the periodic system (Li - Ne) would have a basis set of five functions (two s functions and three p functions).

The most common addition to minimal basis sets is probably the addition of polarization functions, denoted by an asterisk, *. Two asterisks, **, indicate that polarization functions are also added to light atoms (hydrogen and helium). These are auxiliary functions with one additional node. For example, the only basis function located on a hydrogen atom in a minimal basis set would be a function approximating the 1s atomic orbital. When polarization is added to this basis set, a p-function is also added to the basis set. This adds some additional needed flexibility within the basis set, effectively allowing molecular orbitals involving the hydrogen atoms to be more asymmetric about the hydrogen nucleus. This is an important result when considering accurate representations of bonding between atoms, because the very presence of the bonded atom makes the energetic environment of the electrons spherically asymmetric. Similarly, d-type functions can be added to a basis set with valence p orbitals, and f-functions to a basis set with d-type orbitals, and so on. Another, more precise notation

indicates exactly which and how many functions are added to the basis set, such as (p, d).

Another common addition to basis sets is the addition of diffuse functions, denoted in by a plus sign, +, and in Dunning-type sets by "aug" (from "augmented"). Two plus signs indicate that diffuse functions are also added to light atoms (hydrogen and helium). These are very shallow gaussian basis functions, which more accurately represent the "tail" portion of the atomic orbitals, which are distant from the atomic nuclei. These additional basis functions can be important when considering anions and other large, "soft" molecular systems.

The basis set we introduced in the Roothaan-Hall equation will have a radial and an angular part. The angular is almost never commented because it is always the same (s, p, d, f …). It is the radial part which describes how far or how close is the electron with respect to the nucleus. The radial function can be represented by Slater Type Orbitals (STO). However, despite of the adequacy of the STO basis, the integrals are not analytical using STO. Therefore to avoid expensive numerical integrals it is more common to use Gaussian functions (GTO) for which the integrals are usually easier to evaluate. In many cases some modern methods still employ the STO functions. Many textbooks can be found covering the basis set, the different types, their efficiency and the computational requirements(*5*).

1.2.2.6. Correlated Models

Within the Hartree-Fock method of quantum chemistry, the antisymmetric wave function is approximated by a single Slater determinant. Exact wave functions, however, cannot generally be expressed as single determinants. The single-determinant approximation does not take into account Coulomb correlation, leading to a total electronic energy different from the exact solution of the non-relativistic Schrödinger equation within the Born-Oppenheimer approximation. Therefore the Hartree-Fock limit is always above this exact energy. The difference is called the *correlation energy*.

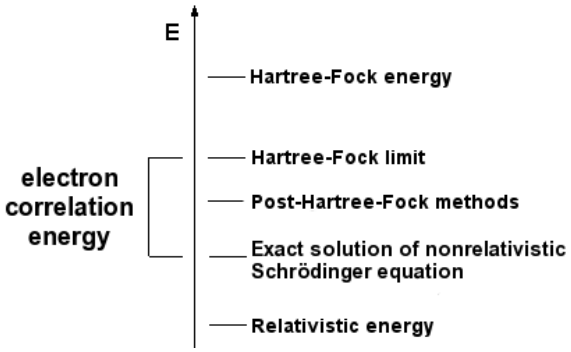


Fig 1.4. Electron correlation energy in terms of various levels of theory of solutions for the Schrödinger equation.

Although many different correlated models have been introduced, here we will focus upon only three classes. Density functional models introduce an "approximate" correlation term in an explicit manner. They offer the advantage of not being significantly more costly than Hartree-Fock models. The "quality" of density functional models obviously depends on the choice of this term, although it is not apparent how to improve on a particular choice. Configuration interaction models and Møller-Plesset models extend the flexibility of Hartree-Fock models by mixing ground-state and excited state wavefunctions. They are significantly more costly than Hartree-Fock models. In the limit of "complete mixing" both configuration interaction and Møller-Plesset models lead to the "exact result", although in practice this limit cannot be reached.

1.2.2.6.1. Kohn-Sham Equation and Density Functional Models

Traditional methods in electronic structure theory, in particular Hartree-Fock theory and its descendants, are based on the complicated many-electron wavefunction. The main objective of density functional theory is to replace the many-body electronic wavefunction with the electronic density as the basic quantity. Whereas the many-body wavefunction is dependent on 3*N* variables, three spatial variables for each of the *N* electrons, the density, $\rho(r)$, is only a function of three variables and is a simpler quantity to deal with both conceptually and practically.

The density functional theory of Hohenberg, Kohn and Sham (*6-9*) is based on the fact that the sum of the exchange and correlation energies of a uniform electron gas can be calculated exactly knowing only its density, $\rho(r)$. In the Kohn-Sham formalism, the ground-state energy, *E*, is written as a sum of the

kinetic energy, E_T, potential energy, E_V, the Coulomb energy, E_J, and the exchange/correlation energy, E_{xc}.

$$E = E_T + E_V + E_J + E_{XC} \tag{1.26}$$

Except for E_T, all components depend on the total electron density, ρ.

$$\rho(r) = 2\sum_{i}^{orbitals} |\Psi_i(r)|^2 \tag{1.27}$$

Here, Ψ_i are so-called Kohn-Sham orbitals and the summation is carried out over pairs of electrons.

1.2.2.6.2 Møller-Plesset Models

Another practical correlation energy scheme is the second-order Møller-Plesset Model, or MP2. This is the simplest number of the class of so-called Møller-Plesset models, the basis of which is the recognition that, while the Hartree Fock wave function and ground-state energy are approximate solutions of the Schrödinger equation, they are exact solution to an analogous problem involving the Hartree-Fock Hamiltonian H_0 in the place of "exact" Hamiltonian, H_{exact}. This involves taking the Hamiltonian of the solvable ***N*** one-electron system, H_0 and perturbing it by λ to give the exact Hamiltonian of the system, such that:

$$H_{exact} = H_0 + \lambda H \tag{1.28}$$

The exact Hamiltonian and energy can be subsequently expanded into a power series:

$$H_{exact} = H_0 + \sum_n \lambda^n H_n \tag{1.29}$$

$$E_{exact} = E_0 + \sum_n \lambda^n E_n \tag{1.30}$$

Where *n* is the number of members in the power series and λ is raised to the power *n*. H_n is the nth order perturbation of H_0 and E_n is the corresponding nth order energy correction for the exact energy. The first-order correction represents the energy associated with the two-electron repulsion integrals. Thus, the HF energy is given by $E+E_1$. The second energy correction, E_2, is the first improvement of the Hartree-Fock energy and is obtained from a *second-order*

Møller-Plesset perturbation calculation (MP2) (*10*). MP2 is size consistent, but doesn't account for electron excitations systematically. Higher-order Møller-Plesset models (MP3, MP4, etc.) have been formulated, but in practice are limited to very small systems. Also, analytical derivatives are not commonly available for these higher-order Møller-Plesset Models, meaning that geometry optimization needs to be done numerically.

1.2.2.6.3 Configuration interaction (CI) Models

This approach involves incorporating different electron configurations into the closed-shell ground-state wavefunction through systematic electron excitations into the ground-state virtual orbitals. This results in a "better" wavefunction that leads to a system with lower energy, hence:

$$\psi_z = c_0\psi_0 + \sum_s c_s\psi_s + \sum_d c_d\psi_d + \sum_t c_t\psi_t + \sum_q c_q\psi_q + \ldots\ldots\ldots \quad (1.31)$$

The better wavefunction, ψ_z, can be given as a linear combination of wavefunctions associated with singlet (ψ_s), doublet (ψ_d), triplet (ψ_t) and higher orders of electron excitation. The hydrogen molecule, for example, has only two electrons; including single and double excitation into the molecular description (using a finite basis set) thus gives full configuration interaction for H_2. CI calculations that involve both single and double excitations are referred to as CISD.

CI is obviously a well defined method and it is variational. However, it is not size-consistent. This can easily be seen by considering the CISD description of a two-electron system, e.g., a helium atom, using just two basis functions. Here, there will be one occupied molecular orbital and one unoccupied molecular orbital, and the CISD description is "exact" (within the confines of this basis set), meaning that all possible electron promotions have been considered. Next, consider the CISD description of two helium atoms at infinite separation. It is not "exact" in that all possible electron promotions have not been considered. Thus, the energies of two helium atoms treated separately and two helium atoms at infinite separation will be different. However, Pople *et al*. have added a quadratic

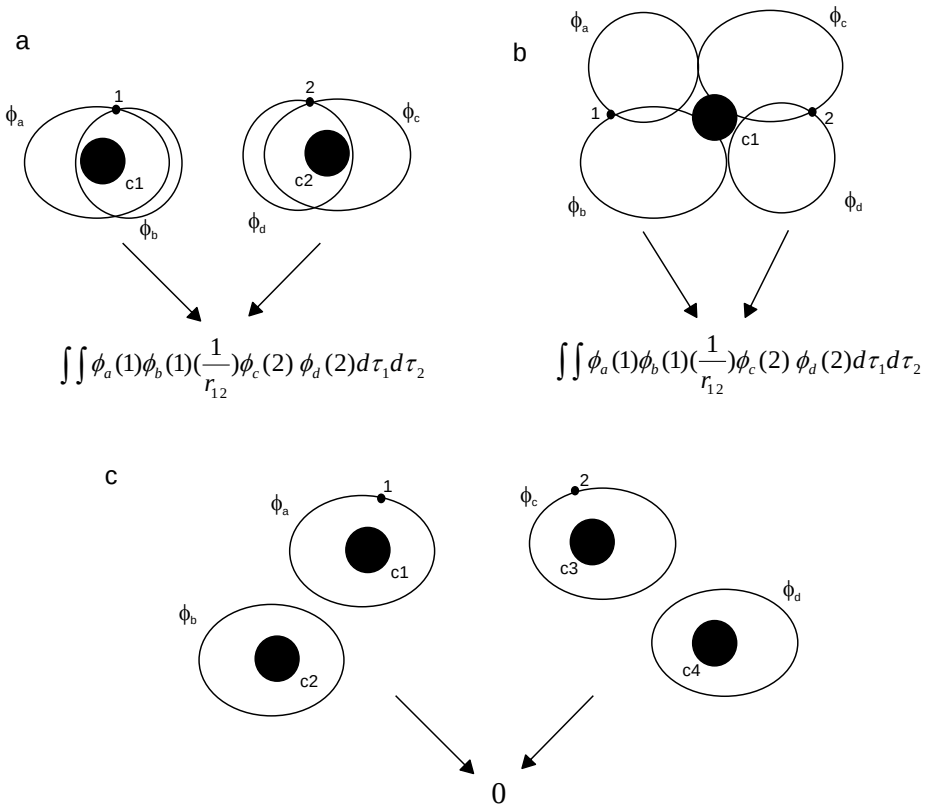
term to the CI method that makes such calculation size-consistent; this version of CI is known as quadratic configuration interaction (QCI) (*11*).

1.2.2.7. The AM1 Semiempirical approximation

The SCF process explained above is a too expensive process for our purpose in enzymatic systems. For both the size and nature of our system and the high number of energy evaluations we will need a cheaper method. The so-called semiempirical methods simplify the HF-SCF equations in such a way that the molecular orbitals and the energy are obtained faster. To consider a less computationally demanding technique must therefore require the calculation of a reduced number of integrals. AM1 (Austin Model 1) (*12*) is based on the Neglect of Diatomic Differential Overlap (NDDO). If two electrons are described by two basis functions, then the two-electron integral is:

$$\int\int \phi_a(1)\phi_b(1)\left(\frac{1}{r_{12}}\right)\phi_c(2)\phi_d(2)d\tau_1 d\tau_2 \qquad (1.32)$$

where basis functions *a* and *b* are used to describe electron 1 and basis functions *c* and *d* describe electron 2. If *a* and *b* are on different centres (Figure 2.5c), then this integral is set to zero under the AM1 technique. This is also the case if *c* and *d* are located on different centres and is in contrast to *ab initio* levels of theory that do calculate these three- and four-centre repulsion integrals. The AM1 method thus evaluates the two-centre two-electron integrals (figure 1.5a), the one-centre two-electron integrals (figure 1.5b), the core-electron attraction integrals and the core-core repulsions. To allow rapid evaluation of these integrals, AM1 is dependent on a set of experimentally determined parameters (e.g. ionisation energies) for each atom type. In addition to calculating a smaller number of integrals, the AM1 method incorporates the 1s electrons and nucleus into a single core. The rationale for this approximation is that important chemical changes involve only the valence electrons.



$$\int\int \phi_a(1)\phi_b(1)\left(\frac{1}{r_{12}}\right)\phi_c(2)\,\phi_d(2)\,d\tau_1 d\tau_2$$

$$\int\int \phi_a(1)\phi_b(1)\left(\frac{1}{r_{12}}\right)\phi_c(2)\,\phi_d(2)\,d\tau_1 d\tau_2$$

Fig 1.5. Electron repulsion integrals are calculated at AM1 for a) two-centre, two-electron and b) one-centre, two-electron interactions. The AM1 method however omits two-electron repulsions over three and four centres, the latter shown in C.

1.2.3. Molecular Mechanics

Molecular mechanics (*2*) describes molecules in terms of "bonded atoms" which have been distorted from some idealized geometry due to non-bonded van der Waals and Coulombic interactions. This is fundamentally different from quantum chemical models, which make no reference whatsoever to chemical bonding. The success of molecular mechanics models depends on a high degree of transferability of geometrical parameters from one molecule to another, as well as predictable dependence of the parameters on atomic hybridization. Even semi-empirical QM methods are currently too computationally demanding to model systems with more than 300 atoms or so. Enzymes and explicit solvation models that consist of many thousands of atoms are therefore unable to be described at this level of theory. Thus molecular mechanics models will be the wisest approach to model these systems. The molecular mechanics "energy" of a molecule is described in terms of a sum of contributions arising from distortion

from "ideal" bond distances ("stretch contribution"), bond angles ("bending contribution") and torsion angles ("torsion contribution"), together with contribution due to "non-bonded" (van der Waals and Coulombic) interactions. It is commonly referred to as "strain energy", meaning that it reflects the "strain" inherent to a "real" molecule relative to some idealized form. This energy function is called an empirical force field and the strategy is the molecular mechanics (MM).

$$E_{struct} = \sum_{bonds} \frac{k_b}{2}(l - l_o)^2 + \sum_{angles} \frac{k_a}{2}(\theta - \theta_o)^2 + \sum_{torsions} \frac{k_t}{2}[1 - cos(n\omega)] + \sum_{impropers} \frac{k_i}{2}(\omega - \omega_o)^2 \quad (1.33)$$

Where l, θ, and ω are the values for bond lengths, angles and torsions respectively. The last term accounts for out-of-plane torsions and is included to maintain planarity about planar centres (e.g. carbonyl systems). The reference values l_o, θ_o, and ω_o are equilibrium geometrical parameters in the absence of any other term in molecular mechanics energy.

In addition to the structural terms, there are also *non-bonded* terms that include electrostatic interactions and van der Waals forces:

$$E_{non\text{–}bond} = \sum_{i=1} \sum_{j=1} \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} + \sum_{i=1} \sum_{j=1} \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}} \right) \quad (1.34)$$

The first term in eqn. 1.34 is the coulombic interaction between two atom-centred partial charges; [12, 13] these are summed over all particles in the system that are not directly bonded to one another. The second term is the van der Waals interaction between species, *i*, and species, *j*, and is often referred to as the *Lennard-Jones 12-6* potential (L-J 12-6). The total energy of the system is usually the summation of the terms in eqn. 1.34 and eqn. 1.33. The majority of computation time associated with MM calculations is due to calculation of the non-bonded interactions.

If every different atom had its own set of parameters the amount of parameters to optimize would be huge. To simplify this task the concept of atom type is used. The atom type idea is based on the common chemical intuition of transferability. For example, all the sp^3 carbon atoms in an alkyl chain will have the same set of

parameters because its chemical behaviour is approximately the same in different length chains. However every force field defines its own set of atoms, so the parameters for a certain atom in a force field are not usually transferable to another force field.

1.2.3.1. Long range interactions

When size increases the most time consuming process is the computation of non-bonding interaction shown in eqn. 1.34. For a pair-wise model the number of non-bonding interactions increase as the square of the number of atoms. However sometimes such computational effort is not necessary, for example, the van der Waals interactions reproduced by the Lennard-Jones potential decays very rapidly.

Non-bonded cutoff:

A very popular strategy is to set a cutoff distance. In this method each atom or group of atoms interacts with all the atoms that are closer than the considered cutoff. The rest of the atoms that fall outside this sphere of interaction do not contribute to the energy nor the gradient of the considered atom. It has been seen that the simple truncation can be too rough; actually in this case the energy is not conserved. Sometimes a cubic switching function is used at a certain distances in order to obtain a smooth decay. At a certain distance r_{on} the switching function is activated so that the potential is reduced until it vanishes at a distance r_{off} . For the switch approximation the following S_{switch} multiples to the potential energy

$$S_{switch}(r)=\begin{cases} 1, & r \le r_{on} \\ \dfrac{(r_{off}^{\,2}-r^2)^2(r_{off}^{\,2}+2r^2-3r_{on}^{\,2})}{(r_{off}^{\,2}-r_{on}^{\,2})^3}, & r_{on} < r \le r_{off} \\ & r > r_{off} \end{cases} \qquad (1.35)$$

Related to the non-bonded cutoff strategy, an additional problem arises. We must compute all the distances between each atom to decide whether its contribution is calculated or not, that is, every atom has a list of the nearest interacting atoms. If we had to calculate this list at every step of a Molecular Dynamics run it would

be too expensive. That is, non-bonded pair-list is built atom-wise or group-wise and it is updated with a certain frequency.

1.2.3.2 Force fields

The AMBER (*13, 14*) CHARMM(*15, 16*) , OPLS (*17*) and GROMOS force fields (*18*) are the most widely used sets of parameters applied to the simulation of biomolecules.

There are studies that perform benchmarks to compare between some of these force fields (*19*) . The first thing that one must realize is that their sets of parameters are different. This is by no means strange because the parameterization process, the atom types and the reference data are different for all of them. So, for example, the atomic partial charges do differ in some magnitude between two force fields. In this sense, we must not give much chemical interpretation to the particular parameters, but they must be seen rather as a set of variables that must be adjusted to give a global good fitting. This is why all these force fields tend to give similar global results. It means for example, that after a Molecular Dynamics simulation some properties such as the conformation distribution are similar in any force field. It is also true that some specific failures of a force field in certain systems have been reported (*16*).

Limitations of molecular mechanics models:

The primary advantage of molecular mechanics models (over any of the quantum chemical methods described earlier in the text) is their simplicity. Molecular mechanics calculations may easily be performed on molecules comprising several thousand atoms. Conformational analysis is perhaps the single most important application of molecular mechanics.

The fact that molecular mechanics models are parameterized may also be considered as an advantage over quantum chemical models. To a great extent, molecular mechanics models could reproduce known experimental data and even anticipate (unknown) data on closely-related systems.

There are important limitations of molecular mechanics models. First, they are limited to the description of equilibrium geometries and equilibrium conformations. Because the MM “strain energy” is specific to a given molecule

(as a measure of how far this molecule deviates from its "ideal arrangement"), strain energies cannot be used in thermochemical calculation.

Second, a molecular mechanics calculation reveals nothing about bonding or, more generally, about electronic distributions in molecules; consequently exploring chemical reactivity and selectivity using molecular mechanics tools is not sensible. There are, however, some situations where purely steric effects are the controlling factor for trends in reactivity and selectivity, and here molecular mechanics would be expected to be of some value.

Third, most currently available force fields are parameterized using results from experiments and that could poorly handle non-equilibrium forms, in particular, reaction transition states.

Finally, due to the fact that molecular mechanics is essentially an interpolation scheme, the success of which depends not only on the choice of parameters, but also on systematics among related molecules.

Molecular mechanics models would not be expected to be highly successful in describing the structures and conformations of "new" (unfamiliar) molecules outside the range of parameterization.

1.2.4. Hybrid methods

Methods for modeling of big systems, such as solute/solvent or in a more general scheme core/environment interaction, can be divided in two big groups. The first one and the earliest is the inclusion of a continuum model characterized by a bulk dielectric constant. These methods are widely used in *ab initio* techniques (*20*) as well as coupled to a Molecular Mechanics force field (*21*), and they provide valuable information. However they do not give specific interactions between the solute and the bulk. In the case of enzymatic reactivity it is this specificity of interaction that makes the enzyme so efficient. Sometimes the system is partitioned at different shells, a core and a first sphere of the environment are modeled explicitly while the outer environment is represented by a continuum model (*22*). Although there are several methods to model solvent effects in biomolecules (*23*) with implicit models they will not be commented further.

The second group is the family of methods where the environment is modeled explicitly at a lower computational cost. These methods are applied successfully

to enzymatic reactions (*24*) , to organometallic catalytic processes (*25*, *26*) and to solid state materials such as zeolites (*27*) .

1.2.4.1. Continuum Solvation Techniques:

While it is usually desirable to incorporate specific solvent-solute interactions within a simulation, especially if a protic-polar solvent is being modelled, computational requirements for such descriptions can be high. An alternative approach is to surround the solute with a solvent description in the form of a continuum with dielectric constant, ε.

The most commonly used continuum solvation method used with *ab initio* levels of theory is the *isodensity polarised continuum model* (IPCM) of Frisch and co-workers (*28*) . With this method, the continuum is modelled using a surface of even electron density that reasonably defines the molecular shape of the solute. The dipole and multipoles of the solute interact with the surrounding dielectric and results in the formation of a dipole in the continuum; this is reflected back on to the solute giving it greater stability. If the electron density of the solute subsequently alters, then there is a corresponding change to the isosurface. The Hamiltonian for a solute immersed in such a continuum usually includes three terms. These are the gas-phase Hamiltonian of the solute; the interactions between the isosurface and the solute electrons; and the interactions between the isosurface and solute nuclei. For the implementation used within this thesis, no geometry optimisation takes place within the continuum. Thus, the calculations give the energy for a solvated *in vacuo* geometry.

1.2.4.2. Quantum Mechanics / Molecular Mechanics (QM/MM) Hybrid Modelling

The earliest hybrid method modeling the environment explicitly is the so-called combined Quantum Mechanics/Molecular Mechanics (QM/MM). In QM/MM (*29-31*) methodology a small reactive part of a chemical system is described by quantum mechanics (QM) whereas the remaining large non-reactive part is described by molecular mechanics (MM).

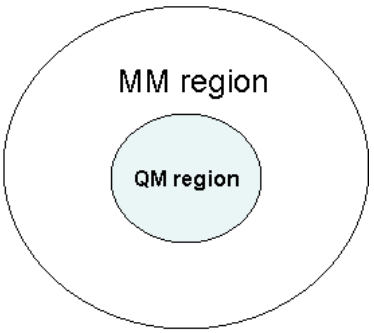


Fig 1.6. The partitioning in a QM/MM system.

There is a significant amount of information reviewing the current status of QM/MM methods in the literature (*32-34*).

The main advantage of QM/MM methods is its easy implementation in computational codes while giving rather good chemical results. Its main disadvantage, especially in enzymatic systems, is to go beyond qualitative results and thus to obtain quantitative numbers out of QM/MM computations. This problem is mainly due to three factors: i) the need for good *ab initio* description for the QM part whereas the usual size of the QM part mostly only allow for semiempirical calculations; ii) the need for accessing free energy numbers through extensive sampling which is too computationally expensive; iii) the difficult calibration of the interaction between the QM part and the MM part, especially in biochemical systems.

Splitting the System:

In most reactive systems, the number of atoms involved in a chemical reaction is fairly limited (*i.e.*, whose electronic properties are changed during the reaction), the rest of the atoms may have a strong influence on the reaction but this is usually limited to short and long range non-bonded interactions that can be represented through both electrostatic and van der Waals interactions. The main idea of QM/MM methodology is to split the chemical system into two parts: the first part is small and described by quantum mechanics, it is called the *quantum part*. The second part is the rest of the system and is described by molecular mechanics, it is called the *classical part*. The full Hamiltonian can therefore be expressed as:

$$H_{eff} = H_{QM} + H_{QM/MM} + H_{MM} \quad (1.36)$$

H_{QM} is the QM Hamiltonian incorporating all sub-atomic particle interactions in the QM part and is given in equation 1.10 for an *N* electron and *M* nuclei system. The QM/MM Hamiltonian represents a cross term in which the QM part is affected by the MM system and is given by:

$$H_{QM/MM} = -\sum_{i=1}^{N}\sum_{m=1}^{X}\frac{Q_m}{r_{im}} + \sum_{\alpha=1}^{M}\sum_{m=1}^{X}\frac{Z_\alpha Q_m}{r_{\alpha m}} + \sum_{\alpha=1}^{M}\sum_{m=1}^{X}\left(\frac{A_{\alpha m}}{r_{\alpha m}^{12}} - \frac{B_{\alpha m}}{r_{\alpha m}^{6}}\right) \qquad (1.37)$$

The first term accounts for the interactions between *N* electrons and *X* MM atoms with point charge, Q_m in a QM/MM system. The term relies upon the co-ordinates of the electrons in the QM region and must therefore be included into the evaluation of the electronic wavefunction. The second term is the interaction between the QM nuclei (with charge *Z*) and the MM atoms (with charge Q_m) separated by distance *r*. The L-J 12-6 potential accounts for the van der Waals interaction between atoms in the MM and QM partitions; it is described by the relevant non-bonded parameters in the MM force field. The inclusion of the L-J potential allows the QM system to differentiate between like-charged, different size species (*cf.* bromide and chloride anions); omission of this term would make any size dependent non-bonded interactions invisible to the QM region. The H_{MM} represents the energy of the MM part of the system.

Dividing Covalent Bonds across the QM and MM Regions:

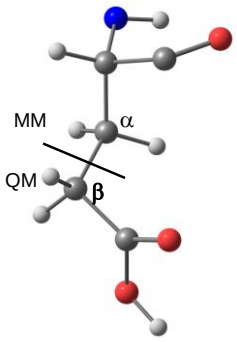


Fig 1.7. Dividing a covalent bond bridges between QM and MM regions in a glutamic acid residue.

In this thesis we have applied a commonly used technique is called the *link atom approximation*. This first involves a decision as to which enzyme residues are

directly involved in the reaction. These residues need to be treated quantum mechanically to allow an electronic description at the reaction centre. However, because the remainder of the enzyme is treated with the simpler MM description, an extra QM atom must be applied to satisfy the valency requirements of the QM system. The quantum link atom used is a hydrogen atom. In figure 1.7, for example, the side-chain of an active site glutamic acid residue is treated quantum mechanically. A QM link atom is placed along the C_α─C_β bond about 1.0Å away from the QM described C_β atom. Thus, the valency of the C_β atom is satisfied and the GLU side chain is modelled by an acetate anion in the QM system. The QM link atom has no non-bonded or internal co-ordinate parameters associated with it and so is not included in the MM part of the calculation. However, the QM link atom is fully visible to other QM atoms and its description is thus included in H_{QM}.

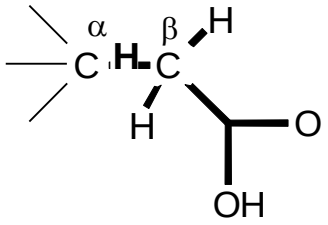


Fig.1.8. QM link atom (H) addition to an enzyme glutamic acid residue between C_α and C_β. This yield "acetic acid" QM system

In the QM/MM system, any MM internal co-ordinate terms (eqn. 1.33) associated with QM atoms alone are deleted. At the QM and MM junction however, such terms for the QM atoms remain and thus allow a full classical description of the environment around the junction MM atom(s).

1.3. Theoretical methods used in this thesis:

Molecular Dynamic Methods

The true picture of a molecular system is far from the static, idealized image provided by molecular mechanics. The atoms and groups in molecules are in constant motion, and, especially in the case of biological macromolecules, these movements are concerted (correlated) and may be essential for biological

function. The direct proof of this behavior is provided by X-ray and NMR experiments. Thermodynamic properties of molecules, especially when they form a complex or are immersed in a solvent, cannot be derived from the harmonic approximation which inherently assumes small amplitude motions around a symmetric minimum. While molecular mechanics and molecular dynamics described below share the same potential energy function they differ conceptually. In molecular mechanics, the information is derived from a single geometry of the molecule. Simulation methods require thousands to millions of geometries to produce meaningful results. The biological action of molecules frequently involves large amplitude motions (biochemists often call them ``conformational transitions'' or ``changes''), resulting in abrupt changes in the geometry of the protein/ligand molecule as well as rearrangement of solvent molecules. The thermodynamic and energetic parameters of such systems can only be derived by realistic and explicit simulation.

There are essentially two approaches to perform such molecular simulations: the stochastic and the deterministic. The stochastic approach is called the Monte Carlo method, but since we have not applied this approach in this thesis, we are not going to get into it.

The deterministic approach, called Molecular Dynamics (MD), actually simulates the time evolution of the molecular system and provides us with the actual trajectory of the system. The information generated from simulation methods can in principle be used to fully characterize the thermodynamic state of the system.

Classical molecular dynamics (MD) can be used to propagate in time the nuclear coordinates of molecular system using the classical equations of motion.

$$-\frac{dV}{dq} = m\frac{d^2q}{dt^2} \tag{1.38}$$

Equation 1.38 must be solved numerically propagating a trajectory at small time-steps. While a typical time-step is about one femtosecond (10^{-15} second) most of chemical interesting events take place at time scales several orders of magnitude higher (micro or millisecond). Therefore the MD equations should be propagated until 10^9-10^{12} steps to observe a reactive event (rare event). Despite of the recent acceleration techniques (*35*) this task is yet too expensive to be performed in the

computers of nowadays This gap of time scales means that so far Molecular Dynamics is rarely used to obtain a real picture of thermally activated chemical processes. Therefore, in this thesis MD will be employed as a technique that may be used to compute equilibrium as well as kinetic properties of a many-body system. In this section, the main topics of molecular dynamics techniques are presented. However it must be taken in consideration that MD is a huge field of research and we will only mention those methods that we will apply to model our enzymatic system. In this sense, a volume of *Accounts of Chemical Research* was dedicated to review the state of the art in Molecular Dynamics simulations of biomolecules (*36*).

1.3.1. Basic equations and algorithms

The position of a set of particles q after a time step displacement Δt can be obtained by a Taylor expansion

$$q_{i+1} = q_i + \frac{dq}{dt}\Delta t + \frac{1}{2}\frac{d^2q}{dt^2}\Delta t^2 + \frac{1}{6}\frac{d^3q}{dt^3}\Delta t^3 \quad (1.39)$$

The second term describes the velocity $V = \frac{dq}{dt}$ and the third term accounts for the acceleration $= \frac{d^2q}{dt^2} = -\frac{1}{m}\frac{dV}{dq}$. Substitution of the velocity and acceleration terms in the equation 1.39, with truncation at third order gives place to the Verlet algorithm which is the basis of the current molecular dynamics simulations techniques

$$q_{i+1} = (2q_i - q_{i-1}) + a_i(\Delta t)^2 \quad (1.40)$$

The acceleration could be taken from the derivatives of the potential energy at the initialization point q_0 and the previous positions could be estimated as $q_{i-1} = q_0 - V_0\Delta t$. One disadvantage of Verlet algorithm is the fact that the velocity does not appear explicitly which is problematic when generating ensembles at constant temperature. The Leap-Frog algorithm overcomes this disadvantage by including the velocity in its equations

$$q_{i+1} = q_i + V_{i+\frac{1}{2}}\,\Delta t \quad (1.41)$$

Where the velocity is

$$V_{i+\frac{1}{2}} = V_{i-\frac{1}{2}} + a_i \Delta t \qquad (1.42)$$

Velocity Verlet or the higher order predictor-corrector is other alternative integration methods. Discussion about their adequacy, the numerical stability, the energy conservation and the time-reversible character must be found in the extensive bibliography (*2*, *37*).

1.3.2. Temperature and Pressure control in MD Simulation

Using the equations of motion (equation 1.38) specified above, the total energy is a constant of motion. In this case the time averages obtained in this MD simulation are equivalent to ensemble averages in *microcanonical ensemble* (NVE). In other words, the number of particles, the volume and the energy of the system are constants. Molecular dynamic simulations within the microcanonical ensembles are the easiest to perform, but it is evident that, if the aim of our simulation is to mimic the conditions under which the systems are investigated experimentally, then the microcanonical ensemble might not be the most appropriate. In particular, it is common to do experiments under conditions in which the ambient temperature and/or pressure are constants. The thermodynamic ensembles that correspond to such conditions are *canonical* or NVT *ensemble*, the *isothermal-isobaric* or NPT *ensemble* and the *isobaric-isoenthalpic* or NPH *ensemble*, respectively.

In this thesis, the NVT ensemble has been adopted. An essential idea behind the method is that the system that we want to simulate is, in fact, not isolated but interacts, or is *coupled*, to an *external bath*. Remember that, in a simulation in the microcanonical ensemble, the energy remains constant and the temperature fluctuates. A coupling to an external system means that the energy can be transferred into and out of the system that we are simulating and so its energy will fluctuates. It is this transfer, properly formulated, which allows the algorithm to control the temperature. The temperature of a particle system is related to the time average of the velocity of the particles

$$\langle \sum_i^n \frac{1}{2} m_i v_i^2 = \frac{3}{2} n k_B T \rangle \qquad (1.43)$$

The initial velocities can be given from a Maxwell-Boltzmann distribution at the desired temperature. And the most intuitive strategy to keep constant the temperature would be to multiply at every step the velocity of the particles by a scaling factor $\lambda = \sqrt{T_{required} \ / T_{current}}$.

A relatively gentle approach is the Berendsen method (*38*).In this approach, the system is coupled to an imaginary external thermal bath held at a fixed temperature T . However, the exchange of thermal energy between the system and the thermal bath is much more gradual. Instead of drastically resetting the velocity of the particle to a new value, the velocity of the particle is gradually scaled by multiplying it by a factor k given by a factor λ

$$\lambda = \left[1 + \frac{\Delta T}{\tau_T}\left(\frac{T}{T_{ins}} - 1\right)\right]^{\frac{1}{2}} \tag{1.44}$$

Where ΔT is the time step and τ_T is the time constant of the coupling. In this way, the velocities of the particles are adjusted such that the instantaneous kinetic temperature T_{ins} approaches the desired temperature T. The strength of the coupling between the system and the thermal bath is controlled through the use of an appropriate coupling time constant τ_T. If rapid temperature control is desired, a small τ_T can be chosen. Consequently, the value of λ will be large and the change in the velocity will be drastic. On the other hand, if weak coupling is needed to minimize the disturbance of the system, a large value can be assigned to τ_T. In the evaluation of their own method, Berendsen *et al*. concluded that static average properties were not significantly affected by the coupling time constant, but the dynamic properties were strongly dependent on it. Their testing showed that reliable dynamic properties could be derived if τ_T was above 0.1 ps. There are a lot of references that well-cover this concept (*39-42*).

1.3.3. Minimum image and periodic boundary

The simulations in condensed phase usually have to reproduce an infinite bulk system such as the solvent that surrounds our solute. If this solvent is reproduced by a single box of solvent the edge of the box will have an artificial surface tension. For situations involving solvent, the small volume of the box in which the macromolecule and solvent are contained introduces undesirable boundary

effects. In fact, the results may depend sometimes more on the size and shape of the box than on the molecules involved. To circumvent this limited box size difficulty, *periodic boundary conditions* are used. This idea is represented in Figure 1.9. In this approach, the original box containing a solute and solvent molecules is surrounded with identical images of itself, i.e., the positions and velocities of corresponding particles in all of the boxes are identical. The common approach is to use a cubic or rectangular parallelepiped box, but other shapes are also possible (e.g., truncated octahedron). By using this approach, we obtain what is in effect an infinite sized system. The particle (usually a solvent molecule) which escapes the box on the right side enters it on the left side, due to periodicity. Since MD simulations are usually performed as an NVE (microcanonical) ensemble (i.e., at constant number of particles, constant volume, and constant total energy) or an NVT (canonical) ensemble, the volume of the boxes does not change during simulation, and the constancy in the number of particles is enforced by the periodicity of the lattice, e.g., a particle leaving the box on left side, enters it on the right side. In the case of periodic boundary conditions we are actually simulating a crystal comprised of boxes with ideally correlated atom movements. Obviously all the possible interactions in an infinite lattice would be infinite. An adequate approximation is the minimum image convention where every atom in the original box interacts with the nearest elements that fall within the given cutoff. These interacting atoms may be in the original box or in the neighboring; however, in order to avoid the interaction of a molecule with itself the cutoff must be smaller than the half of the box.

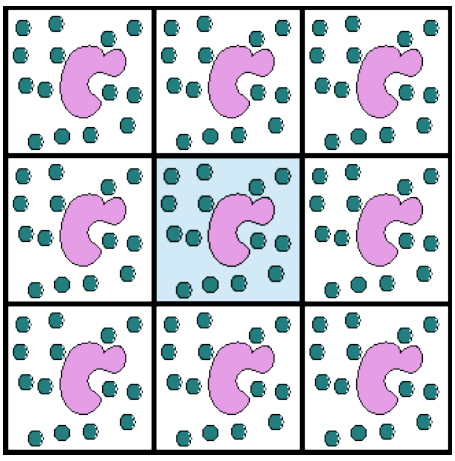


Figure 1.9. Periodic boundary conditions

1.4. Theoretical methods used in this thesis:

Statistical Mechanics

We know for a fact that all matter is composed out of small particles called atoms. Statistical mechanics is that branch of physics which explains the thermodynamic properties of nature starting from a microscopic point of view. Statistical mechanics provides a bridge between the macroscopic realm of classical thermodynamics and the microscopic realm of atoms and molecules. We are able to use computational methods to calculate the thermodynamic parameters of a system by applying statistical mechanics. The statistical mechanics can be used to transform our detailed information at microscopic level obtained during the simulation to thermodynamic magnitudes (*43-45*) .

There are many macroscopic properties that can be calculated from computer simulation of molecular systems *e.g.* heat capacity, radial distribution functions or non-equilibrium properties such as diffusion coefficients. However, here we will only give a short perspective to the free energy calculations. In addition, under the framework of Transition State theory, it will describe the kinetics of such process.

1.4.1. Free Energy Calculations

Free energy is perhaps the most important thermodynamic quantity. Many, if not almost all of the physical properties a chemist or a biochemist can be interested in depend directly or indirectly on the free energy of the system. For example, binding constants, association and dissociation constant and conformational preferences are all directly related to the difference in free energy between alternate states(*46*).

Free energy is a statistical property; it can be seen as a measure of the probability of finding a system in a given state. Furthermore, it is a global property that depends on the extent of the phase (or configuration) space accessible to the molecular system. To get a good estimate of the absolute free energy you would theoretically need to sample the whole phase space, which is not possible.

What can be calculated is the difference in free energy between two related states of a system, which corresponds to the relative probability of finding a system in one state as opposed to the other.

The free energy is usually expressed as the Helmholtz function, *F*, or the Gibbs function, *G*. The Helmholtz function is appropriate for a system with a constant number of particles, temperature and volume (constant NVT; the corresponding ensemble is also referred to as the canonical ensemble) whereas the Gibbs free energy is appropriate for constant number of particles, pressure and temperature (NPT ensemble). While in gas phase and small molecular systems free energy differences can be computed by means of analytical expressions, in condensed phase we have to compute the Potential of Mean Force (PMF) along a distinguished coordinate using sampling techniques.

Before describing the PMF calculation in condensed phase systems, it can be useful to indicate the particular case of free energy calculations in small molecular systems. In the canonical ensemble (NVT) the Helmholtz free energy F is computed from the canonical partition function Z.

$$F(NVT) = -k_B T ln(Z) \tag{1.45}$$

For $\boldsymbol{N}$ independent and identical particles in the classical limit, Z is computed from the partition function of a single molecule z.

$$\mathrm{Z} = \frac{z^N}{N!} \quad ; \; z = \sum_i^{states} e^{E_{tot}\, i/k_B T} \tag{1.46}$$

The total energy of every molecular state E_{tot} can be approximated as a sum involving translational, rotational, vibrational and electronic states (eq. 1.47). This assumption implies that the molecule partition function z_{tot} can be written as a product of terms (eq. 1.48)

$$E_{tot} = E_{trans} + E_{rot} + E_{vib} + E_{elec} \tag{1.47}$$

$$z_{tot} = z_{trans} \cdot z_{rot} \cdot z_{vib} \cdot z_{elec} \tag{1.48}$$

In small molecular systems the partition functions can be approximated by analytical expressions. The term z_{trans} is computed with the free particle model, z_{rot} as the rigid rotor and the z_{vib} is described as a factorization of

normal modes of vibration within the harmonic oscillator. Some improvements exist for anharmonic and hindered rotor (*47*) models but in any case an exploration of the phase space is never performed.

In non-rigid condensed phase systems, the independent particle assumption $\mathrm{Z} = \frac{z^N}{N!}$ is not valid, and the translational partition function z_{trans} cannot be modeled as a free particle. Therefore, we will have to compute the whole partition function through the phase space integral which in the NVT ensemble takes into account all the degrees of freedom of the configuration space.

$$Z = \frac{1}{h^{3N}} \int_q \int_\rho dq\, d\rho\ e^{-\frac{H(q,\rho)}{k_B T}} \tag{1.49}$$

In the Hamiltonian expression, it is possible to perform the integration over the momentum variables, leaving an integral, called the *configuration integral*, of the position coordinates only.

$$H(\rho, q) = \sum_i^n \frac{\rho_i}{2m_i} + V(q) \tag{1.50}$$

Therefore, the integral over the phase space(ρ, q) becomes the configurational integral multiplied by a constant

$$Z \propto \int_q dq\ e^{-\frac{V(q)}{k_B T}} \tag{1.51}$$

Once the partition function and density distribution function are known, the thermodynamic quantities for the system can be determined, for example, the Helmholtz free energy, F (equation 1.45).

1.4.1.1. Potential of Mean Force (PMF)

The Potential of Mean Force is central to many statistical thermodynamical theories. It is, for example, required in some versions of transition-state theory to calculate the rate of reaction between two different states of a system. The PMF of a system with *N* molecules is strictly the potential that gives the average force over all the configurations of all the *n*+1...*N* molecules acting on a particle * at any fixed configuration keeping fixed a set of molecules 1...*n* (*43*).

$$-\nabla_j \omega^{(n)} = \frac{\int e^{-\beta V}\left(-\nabla_j V\right) dq_{n+1} \dots \dots dq_N}{\int e^{-\beta V}\, dq_{n+1} \dots \dots dq_N}$$

$$j = 1,2, \dots \dots, n \qquad (1.52)$$

Where $\beta = 1/k_B\, T$, the $-\nabla_j \omega^{(n)}$ is the average force and therefore $\omega^{(n)}$ is the so-called Potential of Mean Force. PMF can be obtained as a function of one or more of the system's degrees of freedom. This degree of freedom can be a simple function of the Cartesian coordinates of the atoms, such as a distance or an angle, or it can take a more complicated form depending upon the process being studied. A particular example would be $\omega^{(2)}(r_{12})$ that describes the interaction between two molecules held a fixed distance r when the remaining *N-2* molecules are canonically averaged over all configurations. The expression for the PMF is the same as that for a free energy (equation 1.45) except that the averaging is done over all degree of freedom apart from the one corresponding to the variable.

The PMF $\omega(x)$ along some coordinate x is defined from the average distribution function $\langle \rho(x) \rangle$

$$\omega(x) = \omega(x^*) - k_B T ln \left[\frac{\langle \rho\,(x) \rangle}{\langle \rho\,(x^*) \rangle}\right] \qquad (1.53)$$

Where $\omega(x^*)$ and $\langle \rho(x^* \rangle$ are arbitrary reference values. The average distribution function along the coordinate is obtained from a Boltzmann weighted average

$$\langle \rho(x) \rangle = \frac{\int dq \delta\left(x^{'}(q) - x\right) e^{-V(q)/k_B T}}{\int dq\, e^{-V(q)/k_B T}} \qquad (1.54)$$

Where $dq\delta(x^{'}(q) - x)$ is the Dirac delta function for the coordinate. We are assuming that the chosen coordinate x is a geometrical coordinate (q), but x can have any other functionality. For processes with activation barrier higher than the $k_B T$, distribution function $\langle \rho(x) \rangle$ cannot be computed by a straight molecular dynamics simulation. Such computation would not converge due to

low sampling in higher-energy configurations. Special sampling techniques (non-Boltzmann sampling) have been developed to obtain a PMF along a coordinate χ. Although PMF of enzymatic reactions can be calculated using free energy perturbation (*48-50*) , the method used in this thesis and explained here is the Umbrella Sampling technique (*51, 52*).

Umbrella Sampling:

Umbrella Sampling is one way to enhance sampling in a particular region of phase space. It is a means of "bridging the gap" between free energy barriers. The standard Boltzmann weighting for the dynamic sampling is replaced by a potential chosen to cancel the influence of the energy barrier present. In this method, a biasing potential is added to the Hamiltonian to direct the simulations toward a certain goal, for example, to induce a certain conformational transition. The biasing potential is usually in the form of a harmonic potential restraint that keeps the value of a relevant reaction coordinate fluctuating around successive positions along that coordinate. The result of this "stratification" is a series of biased histograms. In order to obtain accurate information about the free energy of the system, the raw data from individual simulations must be unbiased and recombined. The microscopic system is simulated in the presence of an artificial biasing window potential $V_b(x)$

that is added to the potential energy $V(q)$.

$$V' = V(q) + V_b(x) \tag{1.55}$$

Unless the entire range of χ coordinate is spanned in a single simulation, multiple simulations (windows) are performed with different biasing umbrella potentials $V_b(x)_i$ that centre the sampling in different overlapping regions or windows of χ.

The form of the umbrella potential which restricts sampling to a limited to a limited range of values of χ_i is arbitrary. A common choice to produce the biased ensembles is to use for every window i a harmonic function of the form

$$V_b(x)_i = \frac{1}{2}k(x - x_i)^2 \tag{1.56}$$

The superindices b and u indicate *biased* and *unbiased* respectively. The distribution functions from the various windows need to be unbiased $\langle\rho(x)\rangle_i^u$ (the non-Boltzmann factor is removed) and then recombined together to obtain the final estimate PMF $\omega(x)$. Otherwise, obtaining the unbiased PMF for the ith window

$$\omega(x)_i^u = \omega(x^*) - k_B T ln\left[\frac{\langle\rho(x)\rangle_i^b}{\langle\rho(x^*)\rangle}\right] - V_b(x)_i + F_i \qquad (1.57)$$

Where F_i are also undetermined constant that represents the free energy associated with introducing the window potential.

$$e^{-F_i/k_B T} = \langle e^{-V_b(x)_i/k_B T}\rangle \qquad (1.58)$$

The reconstruction of the full distribution function from the separate distribution for each window is the crucial step in an umbrella sampling calculation. An efficient procedure for doing this is the weighted histogram analysis method (WHAM) (*45*). The WHAM method aims to construct an optimal estimate for the average distribution function in the unbiased ensemble $\langle\rho(x)\rangle_i^u$ from the biased distribution functions for each window. The WHAM makes usage of all the information in the umbrella sampling, and does not discard the overlapping regions. In addition, WHAM does not require a significant amount of overlap and it can be easily extended to multi-dimensional PMF (*53, 54*). The WHAM technique computes the total unbiased distribution function $\langle\rho(x)\rangle^u$ as a weighted sum of the unbiased distribution functions $\langle\rho(x)\rangle_i^u$. This weighting function can be expressed in terms of the known biased distribution functions $\langle\rho(x)\rangle_i^b$

$$\langle\rho(x)\rangle^u = \sum_{i=1}^{N} n_i \langle\rho(x)\rangle_i^b \times \left[\sum_{i=1}^{N} n_j\, e^{-(V_b(x)_i - F_i)/k_B T}\right]^{-1} \qquad (1.59)$$

Where N is the number of windows and n_i is the number of data points (number of steps in a MD) in the window. The free energy constants F_i need to be obtained from an optimal estimate of the total distribution function

$$e^{-F_i/k_BT} = \int dx e^{V_b(x)_i/k_BT} \langle \rho(x) \rangle^u \quad (1.60)$$

Equations 1.59 and 1.60 need to be solved iteratively until self-consistence. A common procedure is giving a guess to the set of free energy constants $\{F_i\}$, obtain the total distribution function by equation 1.59 and use it to compute $\{F_i\}$ in equation 1.60. The procedure is stopped when the difference between the constants in two consecutive iterations is below a threshold.

1.4.2. Chemical kinetics

Chemical kinetics, also known as reaction kinetics, is the study of rates of chemical processes. Chemical kinetics includes investigations of how different experimental conditions can influence the speed of a chemical reaction and yield information about the reaction's mechanism and transition states, as well as the construction of mathematical models that can describe the characteristics of a chemical reaction. In this thesis we will focus upon the transition state theory and kinetic isotope effect.

1.4.2.1 Transition state theory

Predicting thermal rate constants of chemical reactions is a major goal in computational chemistry. Such information is required for simulations of complex reacting systems in many areas of science and technologies.

The calculation of rate constants requires a delicate balance between the accuracy of the dynamical theory and the efficiency in obtaining accurate potential energy information. In the extreme of rigorous dynamical treatment, accurate quantum dynamics calculations yield detailed state-to-state reactive cross sections or rate constants with full consideration of quantum effects. However, such calculations are currently limited to systems having very small number of atoms with the use of global analytical potential energy functions. At the other extreme, transition state theory (TST) has been practical for a wide range of chemical processes due to its simplicity. The basic model only requires

potential energy information at the reactant(s) and transition state, since it treats many dynamical effects only approximately. Thus, such information can be obtained from accurate electronic structure calculations. There are, however, many reactions in which variational effects and/or tunneling are important and for these systems more accurate dynamical treatments are desirable.

Transition state theory (TST) is one of most successful theories in theoretical chemistry (*55*). It gives the framework of chemical reaction rate theory and today it is the general name for many theories based in whole or in part on the fundamental assumption of the existence of a hypersurface (transition state) in the phase space that divides reactants and products. In the transition state theory, the mechanism of interaction of reactants is not considered; the important criterion is that colliding molecules must have sufficient energy to overcome a potential energy barrier (the activation energy) to react.

Transition state theory or activated complex theory provides a simple formalism for obtaining thermal rate constant by mixing the important features of the potential energy surface with a statistical representation of the dynamics. In addition to the Born- Oppenheimer approximation, TST is based on three assumptions:

1- Classically there exists a surface in phase space that divides it into a reactant region and a product region. It is assumed that this dividing surface is located at the transition state, which is defined as the maximum value on the minimum energy path (MEP) of the potential energy surface that connects the reactant(s) and product(s). Any trajectory passing through the dividing surface (or bottleneck) from the reactant side is assumed to eventually form products. This is often referred to as the non recrossing rule.

2- The reactant equilibrium is assumed to maintain a Boltzmann energy distribution.

3- Activated complexes are assumed to have Boltzmann energy distributions corresponding to the temperature of the reacting system. These activated complexes are defined as super-molecules having configurations located in the vicinity of the transition state.

The rate constant at a given temperature T is

$$k(T) = \gamma(T)\frac{1}{\beta h}e^{-\beta \Delta G^{\neq}(T)} \quad (1.61)$$

The magnitude $\Delta G^{\neq}(T)$ is the free energy difference between the reactants and the transition state. The transmission coefficient $\gamma(T)$ is a correction term that stands for all the approximations assumed in the TST.

$$\gamma(T) = g(T)\kappa(T)\Gamma(T) \quad (1.62)$$

Where $g(T)$ accounts for the deviations from equilibrium distribution in phase space. It can be less than or greater than 1. $\Gamma(T)$ arises from dynamical recrossing. It will be 1 or less than 1. $\kappa(T)$ is the contribution from quantum mechanical tunneling therefore almost always this correction is greater or equal to 1.

Further information regarding the current status of the TST (*56*) and its application to enzymatic systems (*25*) may be found in recent reviews.

However, the most crucial task is the computation of the free energy difference $\Delta G^{\neq}(T)$, which must be calculated along a predefined reaction coordinate. In most simple cases, such as proton transfers, and therefore the reaction coordinate is adequately chosen. On the other hand, many other reactions have a complicated or unpredictable mechanism and therefore the choice of a predefined reaction coordinate is quite tricky.

1.4.2.2. Isotope Effects

Theoretically modelling transition structures can lead to the calculation of experimentally observable properties, namely, the kinetic and equilibrium isotope effects (KIE and EIE respectively) (*57*). Such calculations can therefore allow full interplay between theory and experiment, in which theory is able to give insight into the experimentally non-determinable TS structure.

Some basic understanding:

The **kinetic isotope effect** (**KIE**) is a variation in the reaction rate of a chemical reaction when an atom in one of the reactants is replaced by one of its isotopes. A KIE involving hydrogen and deuterium is represented as:

$$KIE = \frac{k_H}{k_D} \tag{1.63}$$

Where k_H and k_D are reaction rate constants for hydrogen and deuterium, respectively (figure 1.10).

The fundamental vibrational frequency υ of a chemical bond between atom A and B is, when approximated by a harmonic oscillator

$$\upsilon = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \tag{1.64}$$

Where k is the spring constant for the bond and μ is the reduced mass of the A-B system:

$$\mu = \frac{m_A m_B}{m_A + m_B} \tag{1.65}$$

Quantum mechanically, the energy of the *n*-th level of a harmonic oscillator is given by:

$$E_n = h\upsilon\left(n + \frac{1}{2}\right) \tag{1.66}$$

Thus, the zero-point energy (n = 0) will decrease as the reduced mass increases. With lower zero-point energy, more energy is needed to overcome the activation energy for bond cleavage.

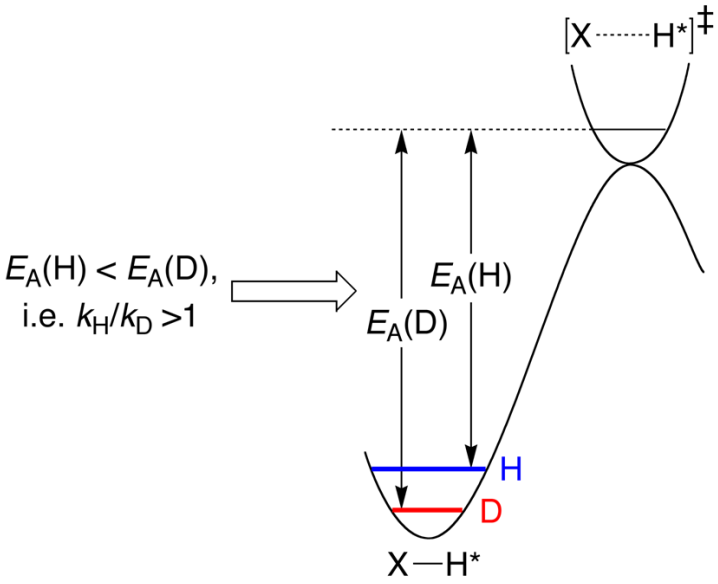


Fig. 1.10. The origin of primary kinetic isotope effect using X-H/X-D as example

Because a neutron has a unit charge of zero, its subsequent addition to a nucleus doesn't alter the electronic description of the system under the Born-Oppenheimer approximation. Such a change thus results in no alteration of the PES or its curvature. The force constants contained in the hessian therefore have the same value for an isotopically substituted system as compared with non-isotopically substituted case. However, the vibrational frequencies are altered because the mass (and hence the reduced mass) of the molecule is increased. The zero-point vibrational energy for a particular vibration (e.g. bond stretch) is reduced with a heavier isotope; for an isotopically substituted system therefore, a greater amount of energy is required to dissociate the molecule.

Boltzmann distribution law describes the individual populations of i energy levels for a particular system vibration

$$\frac{n_i}{n_o} = exp - \{(\varepsilon_i - \varepsilon_o)/k_B T\} \quad (1.67)$$

n_i is the number of molecules found in energy level i with n_o being the population of the ground vibrational state. The corresponding energies for each energy level are given by ε_i and ε_o. It follows that the summation of all the individual energy level populations Σn_i can be related to the total number of particles in the system by

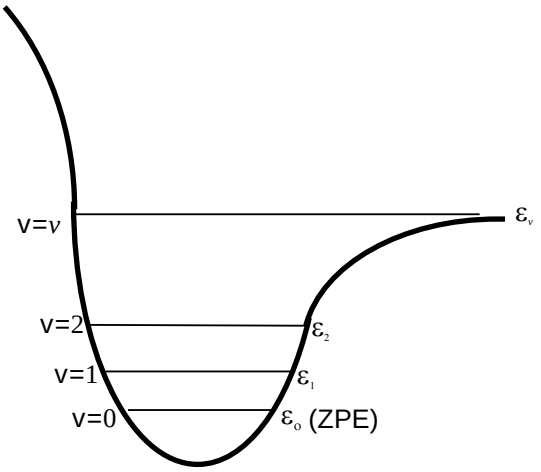


Fig. 1.11. Potential curve for one degree of freedom within a molecular system indicating the quantized energy levels.

$$N = n_o \sum_{i=0}^{n} exp - [(\varepsilon_i - \varepsilon_o)/k_B T] \quad (1.68)$$

The summation is known as the partition function Q and represents the distribution of N species within v energy levels for a particular degree of freedom (i.e.$N = n_o Q$).

For two species in equilibrium in which the ZPE levels are separated by $\Delta\varepsilon_o$ the equilibrium constant is:

$$K = \frac{N_a}{N_b} = \frac{Q_a}{Q_b} . exp - [\Delta\varepsilon_o / k_B T] \quad (1.69)$$

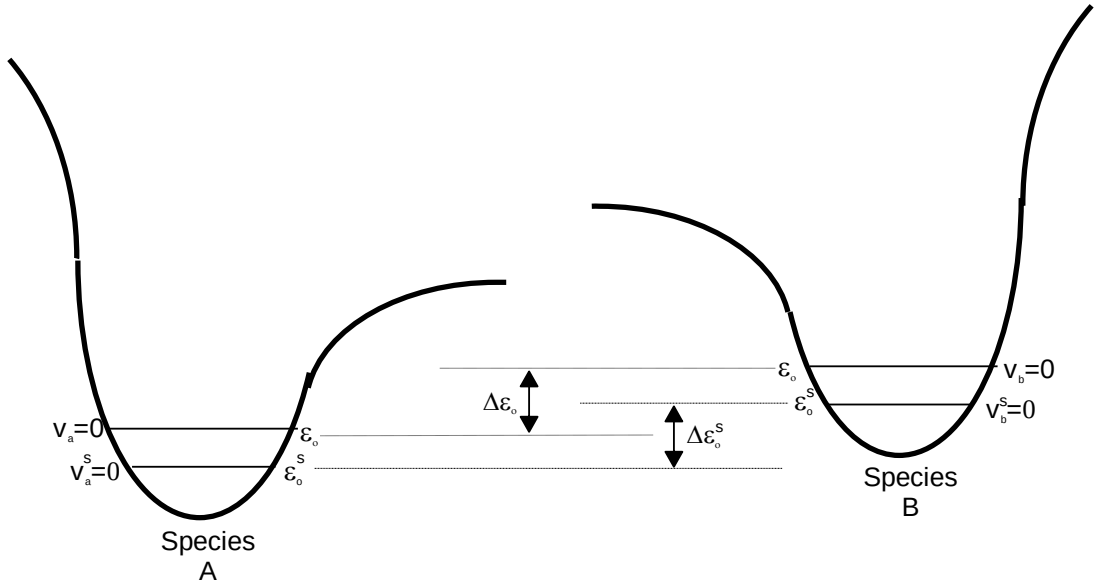


Fig. 1.12. Schematic diagram showing the differences in the ZPE along the "reaction co-ordinate" degree of freedom for species A and B in equilibrium

The resulting energy difference between the ZPE levels for isotopically substituted *A* and *B* species is likely to be different as compared to the unsubstituted molecules. This leads to a different the equilibrium constant K^s for the substituted system:

$$K^s = \frac{N_a{}^s}{N_b{}^s} = \frac{Q_a{}^s}{Q_b{}^s}.exp - [\Delta\varepsilon_o{}^s/k_BT] \qquad (1.70)$$

The resulting *equilibrium isotope effect* is presented as a K/K^s fraction and is given by:

$$\frac{K}{K^s} = \frac{Q_a}{Q_b}.\frac{Q_b{}^s}{Q_a{}^s}.exp - [(\Delta\varepsilon_o - \Delta\varepsilon_o{}^s)/k_BT] \qquad (1.71)$$

where $\Delta\varepsilon_o$ and $\Delta\varepsilon_o{}^s$ are defined in figure 1.11. A similar equation can be derived for the *kinetic isotope effect* (KIE) except state *B* is now located at a first-order saddle point (and thus labelled *TS*). The resulting rate constant ratio is given by:

$$\frac{K}{K^s} = \frac{Q_a}{Q_{TS}}.\frac{Q_{TS}{}^s}{Q_a{}^s}.exp - [(\Delta\varepsilon_o - \Delta\varepsilon_o{}^s)/k_BT] \qquad (1.72)$$

Where $\Delta\varepsilon_o$ and $\Delta\varepsilon_o{}^s$ are now zero-point vibrational activation energies for the unsubstituted and substituted processes respectively and are shown in figure 1.12

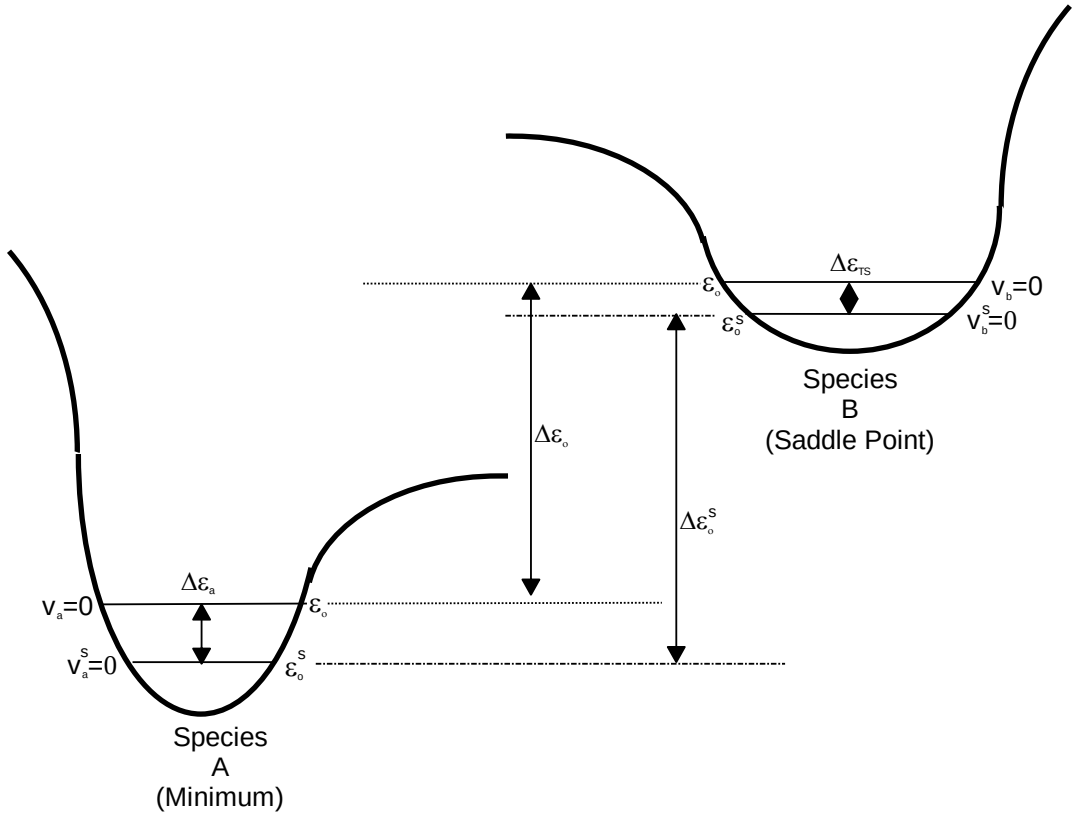


Fig.1.13 Schematic diagram showing the differences in the ZPE along the "reaction co-ordinate" degree of freedom for species A and B, the latter being located at a first-order saddle point on the PES

An alternative expression for eqn. 1.72 and eqn. 1.71 can be given in terms of the energy gaps between the substituted and unsubstituted molecule for both states under study. For example, the rate constant ratio of eqn 1.73 can be given as:

$$\frac{K}{K^s} = \frac{Q_a}{Q_{TS}} . \frac{{Q_{TS}}^s}{{Q_a}^s} . exp - [({\Delta\varepsilon_o}^{TS} - {\Delta\varepsilon_o}^a)/k_B T] \qquad (1.73)$$

Here, $\Delta\varepsilon_o{}^a$ represents the energy gap between the zero vibrational energy level for the substituted and unsubstituted molecule, *A*, and $\Delta\varepsilon_o{}^{TS}$ is the corresponding difference for the studied TS (figure 1.13). If the energy gap associated with state *A* is larger than that for the TS, then the exponential term is positive and the resulting ratio is more than one. This is known as a *normal isotope effect*; it

occurs when the potential of the system becomes looser on moving from reactants to transition-state.

Conversely, if the energy gap associated with the TS is larger, then a negative exponential term is obtained and the rate ratio has a value of less than one. This is known as an *inverse isotope effect*; it occurs when the potential becomes stiffer on moving from reactants to transition-state.

Primary kinetic isotope effects (1^o KIE) involve isotopic substitutions on atoms directly associated with bond making or bond breaking. These effects may have values of about 7.0 for primary deuterium effects, while heavy atom substitutions result in much smaller effects (*ca.* 1.05). *Secondary α-deuterium effects* involve isotopic substitutions on hydrogen atoms bonded to the reacting centre. These atoms are not directly involved with the bond making or bond breaking process and a smaller isotope effect is likely to result compared to primary effects. The rate ratio values obtained for secondary α-deuterium effects tend to be in the range of 0.80-1.25.

More appropriate substitutions as regards the studies presented in this thesis are *secondary β-deuterium effects* (2^o-β-KIE). These result from isotopic substitution involving hydrogens on the adjacent site to the reaction centre. Their effect can be given

$$ln\,(k_H/k_D) = cos^2\theta ln\,(k_H/k_D)_{max} + (k_H/k_D)_i \qquad (1.74)$$

The first term on the right hand side of equation 1.74 accounts for the hyperconjugation effect obtained from the dihedral alignment between one of the β-hydrons and the developing empty p-orbital at the reacting centre. The empty p-orbital interacts with the anti-bonding orbital of the β-hydron, thus loosening the bond to the hydrogen while double bond character is gained along the *adjacent centre − reactant site* bond. The maximum hyperconjugative effect $(k_H/k_D)_{max}$ is obtained when there is maximum overlap between the empty p-orbital and a β-hydrogen (i.e. when $\theta = 0^o$).

The second term is a slight electron donating effect associated with the deuterium system (this opposes hyperconjugation). This inductive effect is usually minor compared to the hyperconjugation also taking place.

Finally, a *solvent isotope effect* is obtained when H_2O solvent is replaced with D_2O. This may be due to a primary effect in which solvent participates explicitly in the rate-determining step of a reaction. Additionally, general base catalysis can occur in which a proton is transferred from one water molecule to the other. Furthermore, a secondary effect associated with non-cleaved solvent protons may also exist. The solvent isotope effect therefore, is a complex combination of effects involving many solvent molecules.

1.4.2.2.1. Kinetics of Enzyme Catalyzed Reactions (*58*)

Michaelis and Menten provided mathematical treatment for the kinetics of an enzyme catalysed reaction involving the conversion of substrate, S, into product, P in the presence of enzyme, E (*59*).

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

In this scheme, it is assumed that the substrate first binds to the enzyme to form an enzyme-substrate complex $[ES]$ before forming the product. It is also assumed that the second stage is non-reversible and that only one ES complex is formed during the reaction. The $[ES]$ concentration can be given in terms of the measurable starting concentration of the enzyme $[E]_o$ provided that $[ES]$is constant during the reaction. Using steady state kinetics:

$$\frac{d[ES]}{dt} = k_1[E]_o - [ES][S] - k_{-1}[ES] - k_2[ES] = 0 \qquad (1.76)$$

From the steady state assumption, an expression can be formulated for $[ES]$. Using equation 1.76

$$[ES](k_1[S] + k_{-1} + k_2) = k_1[E]_o[S] \qquad (1.77)$$

Rearranging and dividing through by k_1 results in an expression for [ES]:

$$[ES] = \frac{[E]_o[S]}{\frac{k_{-1}+k_2}{k_1}+[S]} \tag{1.78}$$

The rate constant fraction on the denominator of equation 1.78 is known as the Michaelis constant, K_m. It represents a pseudo-equilibrium constant for the formation and destruction of the ES complex over the whole reaction scheme. The overall rate of reaction is then given by $k_2[ES]$:

$$rate = \frac{V_{max}\,[S]}{K_m + [S]} \tag{1.79}$$

Where $V_{max} = k_2[E]_o$ and represents the limiting rate for the enzyme catalysed reaction. If the substrate concentration is in excess, then the value of K_m is insignificant in comparison, and the rate is approximately equal to the limiting rate for the reaction. However, if the substrate is under limiting concentrations, then the value of K_m is more significant than $[S]$ on the denominator. These situations can be summarised as:

if $[S] \gg K_m$ $rate \approx V_{max}$

if $[S] \ll K_m$ $rate \approx \frac{V_{max}\,[S]}{K_m}$

There exists therefore, two conditions under which isotope effects can be evaluated. When $[S]$ is in excess, it is assumed that all the enzyme is complexed and the rate determining step may be a single or a series of chemical changes involving the ES complex. However, under limiting concentrations, the rate will also be partially dependent on substrate binding to the enzyme. Any V_{max}/K_m isotope effect must therefore not only include the chemical step(s) involving the ES complex, but also the EIE for the binding of the substrate to the enzyme as well.

References

1 Nye, M. J., *From Chemical Philosophy to Theoretical Chemistry: Dynamic of Matter and Dynamic of Disciplines*, 1800-1950: University of California press, 1994.

2 Leach, A.R., *Molecular modelling: principles and applications*, Longman, Harlow, 1996.

3 Hinchliffe, A., *Modelling Molecular Structures*, Wiley, Chichester, 1996.

4 Atkins, P. W., and Friedman, R. S., *Molecular Quantum Mechanics (3rd Ed.)*, Oxford University Press, Oxford, 1997.

5 Jensen, F., *Introduction to Computational Chemistry*, Wiley, Chichester, 1999.

6 Hohenberg, P., and Kohn, W., (1964) Inhomogeneous electron gas, *Phys. Rev*., **136**, B864-871.

7 Kohn, W., and Sham, L. J., (1965) Self-consistent equations including exchange and correlation effects. *Phys. Rev*., **140**, A1133-1138.

8 Parr, R. G., and Yang, W., *Density Functional Theory of Atoms and Molecules*, Oxford University Press, Oxford, 1988.

9 Kohanoff, J., *Electronic Structure Calculations for Solids and Molecules: Theory and Computational Methods*, Cambridge University Press, 2006.

10 Foresman, H. B., and Friedman, R. S., *Molecular Quantum Mechanics (3rd Ed.)*. Oxford University Press, Oxford, 1997.

11 Pople, J. A., Head-Gordon, M., and Raghavachari, K. (1987) Quadratic configuration. A general technique for determining electron correlation energies. *J. Chem. Phys.*, **87**, 5968-5975.

12 Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. (1985) AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, **107**, 3902-3909.

13 Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. (1984) A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins, *Journal of the American Chemical Society 106*, 765-784.

14 Pearlman, D., Case, D., Caldwell, J., Ross, W., Cheatham, T., Debolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) AMBER, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular- dynamics and free-energy calculations to simulate the structural and energetic properties of molecules, *Comp. Phys. Comm.* **91**, 1-41.

15 Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) Charmm: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187-217.

16 MacKerell Jr., A. D., Bashford, D., Bellott, M., Dunbrack Jr., R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher III, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem.* **102**, 3586-3616.

17 Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc.* **118**, 11225-11236.

18 Scott, W. R. P., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Krüger, P., and van Gunsteren, W. F. (1999) The GROMOS biomolecular simulation program package, *J. Phys. Chem.* **103**, 3596-3607.

19 Price, D. J., and Brooks III, C. L. (2002) Modern protein force fields behave comparably in molecular dynamics simulations. *J. Comput. Chem.* **23**, 1045-1057, 2002.

20 Cramer, C. J., and Truhlar, D. G. (1999) Implicit solvation models: equilibria, structure, spectra, and dynamics, *Chem. Rev.* **99**, 2161-2200.

21 Roux, B. *Computational biochemistry & biophysics*, Marcel Dekker Inc., New York, 2001.

22 Cui, Q. (2002) Combining implicit solvation models with hybrid quantum mechanical/molecular mechanical methods: a critical test with glycine, *J. Chem. Phys.* **117**, 4720-4728.

23 Orozco, M., and Luque, F. J. (2000) Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.* **100**, 4187-4225.

24 Garcia-Viloca, M., Gao, J., Karplus, M., and Truhlar, D. G., (2004) How enzymes work:analysis by modern rate theory and computer simulations. *Science,* **303**, 186-195.

25 Matsubara, T., Maseras, F., Koga, N., and Morokuma, K. (1996) Application of the new "integrated MO + MM" (IMOMM) method to the organometallic reaction Pt (Pr) + H (R = H, Me, t-Bu, and Ph), J. *Phys. Chem.* **100**, 2573-2580.

26 Maseras, F., and Lledós, A. *Computational modeling of homogeneous catalysis*. Kluwer, Dordrecht (Holland), 2002.

27 Sierka, M., and Sauer, J. (2000) Finding transition structures in extended systems: A strategy based on a combined quantum mechanics empirical valence bond approach, *J. Chem. Phys.* **112**, 6983-6996.

28 Foresman, J. B., Keith, T. A., Wiberg, K. B., Snoonian, J., and Frisch, M. J. (1996). Solvent Effects. 5. Influence of cavity shape, truncation of electrostatics, and electron correlation on ab initio reaction field calculations. *J. Phys. Chem.*, **100**, 6098-16104.

29 Warshel, A., and Levitt, M. (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol*,. **103**, 227-249.

30 Singh, U. C., and Kollman, P. A. (1986) A combined *ab initio* quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.*, **7**, 718-730.

31 Field, M. J., Bash, P. A., and Karplus, M. (1990) A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **11**, 700-733.

32 Gao, J., and Thompson, M. A., eds. *Combined Quantum Mechanical and Molecular Mechanical Methods*. ACS Symposium Series 712 Washington D.C.: American Chemical Society.1998.

33 Monard, G., and Merz, K. M. (1999) Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Acc. Chem. Res.* **32**, 904-911.

34 Field, M. J. (2002) Simulating enzyme reactions: challenges and perspectives. *J. Comput. Chem.* **23**, 48-58.

35 Elber, R., Ghosh, A., and Cárdenas, A. (2002) Long time dynamics of complex systems. *Acc. Chem. Res.* **35**, 396-403.

36 Issue dedicated to molecular dynamics simulations of biomolecules. (2002) *Acc. Chem. Res.* **35** (issue 6).

37 Allen, M. P., and Tildesley, D. J. *Computer Simulation of Liquids*, Oxford University Press, Oxford, 1987.

38 Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690.

39 Nosé, S. (1984) A unified formulation of the constant temperature molecular dynamics methods, *J. Chem. Phys.*, **81**, 511-519.

40 Hoover, W. G. (1985) Canonical dynamics: Equilibrium phase-space distributions, *Phys. Rev. A* **31**, 1695-1697.

41 Martyna, G. J., and Klein M. L. (1992) Nosé-Hoover chains: The canonical ensemble via continuous dynamics, *J. Chem. Phys.* **97**, 2635-2643.

42 Cheng, A., and Merz Jr., K. M. (1996) Application of the Nosé-Hoover chain algorithm to the study of protein dynamics, *J. Phys. Chem.* **100**, 1927-1937.

43 McQuarrie, D. A. *Statistical Mechanics*. University Science Books, Sausalito, California, 2000.

44 Chandler, D. *Introduction to Modern Statistical Mechanics*, Oxford University Press. New York 1987.

45 Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J. M. (1992) The weighted histogram analysis method for

free-energy calculations on biomolecules, *J. Comput. Chem.* **13**, 1011-1021.

46 Kollman, P. (1993) Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395-2417.

47 Masgrau, L., González-Lafont, À. and Lluch, J. M. (2003) Dependence of the rate constants on the treatment of internal rotation modes, *J. Comput. Chem.* **24**, 701-706.

48 Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, New York, 1992.

49 Li, G., Zhang, X., and Cui, Q. (2003) Free energy perturbation calculations with combined QM/MM potentials complications, simplifications, and applications to redox potential calculations, *J. Phys. Chem. B,* **107**, 8643-8653.

50 Stanton, R. V., Dixon, S. L., and Merz, Jr., K. M. (1997) Free energy perturbation calculations within quantum mechanical methodologies. In: *Computational Approaches to Biochemical reactivity*, Naray-Szabo, G., Warshel, A., eds., Kluwer Academic Publishers Dordrecht, 103-123, 1997.

51 Torrie, G. M., and Valleau, J. P. (1997) Nonphysical sampling distributions in Monte Carlo free energy estimation: umbrella sampling, *J. Comput. Phys.* **23**, 187-199.

52 González-Lafont, A., Lluch, J. M., and Bertrán, J. Monte Carlo simulations of chemical reactions in solution. In: *Solvent Effects and Chemical Reactivity: Understanding Chemical Reactivity*, Tapia, O., Bertrán, J., eds., Kluwer Academic Publishers Dordrecht, 125-177, 1996.

53 Boczko, E. M., and Brooks III, C. L. (1993) Constant-temperature free energy surfaces for physical and chemical processes, *J. Phys. Chem.* **97**, 4509-4513.

54 Roux, B. (1995) The calculation of the potential of mean force using computer simulations, *Comput. Phys. Commun.* **91**, 275-282.

55 Eyring, H. (1935) The activated complex and the absolute rate of chemical reactions, *Chem. Rev.* **17**, 65-77.

56 Truhlar, D., Garrett, B., and Klippenstein, S. (1996) Current status of transition-state theory, *J. Phys. Chem.* **100**, 12771-12800.

57 Maskill, H., *The physical basis of organic chemistry*. Oxford University Press, Oxford, 1985.

58 Cornish-Bowden, A., *Fundamentals of Enzyme Kinetics (Rev. Ed.).* Portland Press, London, 1995.

59 Michaelis, L., and Menten, M. L. (1913). *Biochem. Z.*, **49**, 333-369.

2. Glycosidase Enzymes

Chapter overview

Since glycosidase enzymes have already been an area of much interest in Prof. Ian Williams' research group, in this chapter I give a general overview on glycosidases, summarizing many topics, specifically those that are relevant to computational chemists in general and to our research group in particular.

2.1 Carbohydrates

Carbohydrates, in the form of glycoproteins, glycolipids and polysaccharides, are fundamental to a wide range of biological processes. The formation and degradation of the polymers are done via the glycosidic bond, which is one of the most stable linkages found within natural biopolymers. The stereochemical variation in assemblies of monosaccharides is enormous, both in the type and manner of the linkage units. The stability and the wide disparity in structure and function require an extensive list of efficient enzymes with high specificity for the various biological systems. The enzymes responsible for this task comprise the glycosidases (also called glycoside hydrolases, GH), glycotransferases and phosphorylases. Glycosidases hydrolyze the glycosidic bond, the glycotransferases perform glycoside bond formation, and the phosphorylases either break the glycosidic bond via phosphorolysis or polymerize the oligosaccharide substrates with a sugar phosphate as donor substrate. As glycosidases are the focus of this work, the other two groups will not be described in more detail.

2.2. Families and folds

A particularly valuable contribution to the understanding of glycosidases has been their classification into families on the basis of sequence similarities. A classification of glycosyl hydrolases into more than 45 families, on the basis of similarities in amino acid sequence, has been proposed (*1*, *2*), with the prospect

that this may facilitate the derivation of useful information on the structure and function of these enzymes. Underlying this classification was the idea that proteins in a given family would have a fold sufficiently similar to allow homology modelling. In this classification, enzymes with different substrate specificities are sometimes found in the same family, indicating an evolutionary divergence to acquire new specificities, as found in, for example, families 1, 13 and 16. On the other hand, enzymes that hydrolyze the same substrate are sometimes found in different families.

Because the 3D structures of proteins are more highly conserved than their sequences, several sequence-based families may have related folds. For instance, a structural similarity was suggested for family 11 xylanases and family 12 cellulases (*3*), whereas family 7 cellulases have been found to have an arrangement of catalytic residues and a fold similar to those of the β-1,3-glucanase and β-1,3-1,4-glucanase of family 6 (*4*). More recently families 1, 2, 5, 10, 17, 30, 35, 39 and 42 were proposed to have evolved from a common ancestor (*5, 6*).

Active sites have frequently located within these structures through complexes with ligands or from the location of residues identified as important to catalysis by labeling studies.

2.3. Sugar subsite nomenclature

The nomenclature used for sugar-binding subsites in glycosidases indicates the the cleavage point by naming the subsite on the reducing end of the sugar +n and the subsite on the non-reducing end –n, where n is an integer. The cleavage occurs between subsite -1 and +1 (Fig. 2.1) (*7*)

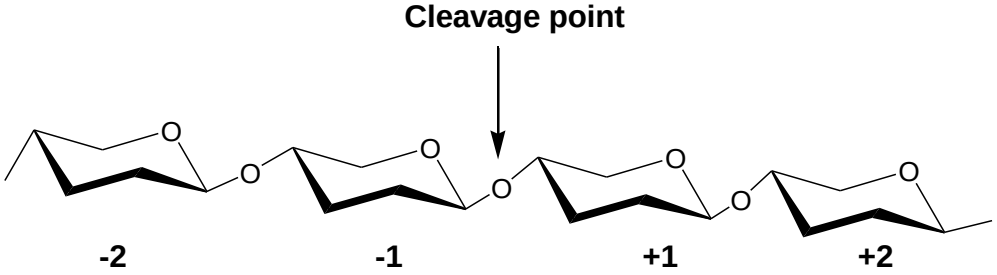


Fig.2.1. Nomenclature of sugar-binding sites. The cleavage occurs between subsite -1 and +1.

2.4. Active-site topologies

Although many protein folds are represented in the 22 families for which a 3D structure is known, the overall topologies of the active sites fall into only three general classes (*8*). Figures, 2.1 and 2.2 are taken from reference 7.

Pocket or crater

This sort of topology (Fig. 2.2(i)a) is optimal for the recognition of a saccharide non-reducing extremity and is encountered in monosaccharidases such as β-galactosidase, β-gluconase, sialidase (or neuraminidase), and in exopolysaccharidases such as glucoamylase and β-amylase. On the other hand, these enzymes are not very efficient for fibrous substrates such as native cellulose, which has almost no free chain ends.

Cleft or groove

This is an open structure of the enzyme (Fig. 2.2(i)b) in which several sugar units could bind in a random manner and that sort of binding is commonly found in endo-acting polysaccharidases such as lysozymes, endocellulases, chitinases, α-amylases, xylanases, and β-1,3-1,4-glucanases

Tunnel

This topology (Fig. 2.2(i)c) arises from the previous one when the protein evolves long loops that cover part of the cleft. This topology allows these enzymes to release the product while remaining firmly bound to the polysaccharide chain, thereby creating the conditions for processivity (Fig. 2.2(ii)). It remains unclear the fashion by which the substrate penetrates the active site but it should be noted that, depending on the mechanism (inverting or retaining) and the exact position of the cleavage point, the directionality of the enzyme motion along the chain may change.

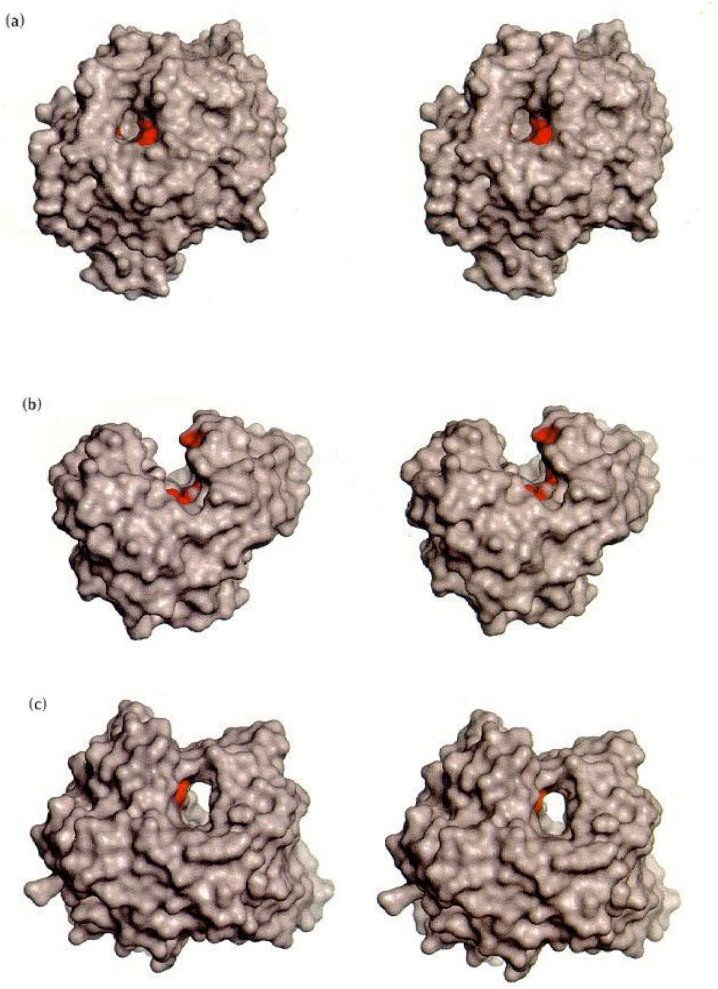


Fig. 2.2(i). The three types of active site found in glycosyl hydrolases. **(a)** The pocket (glucoamylase from *A. awamori*). **(b)** The cleft (endoglucanase E2 from *T. fusca*). **(c)** The tunnel (cellobiohydrolase II from *T. reesei*).

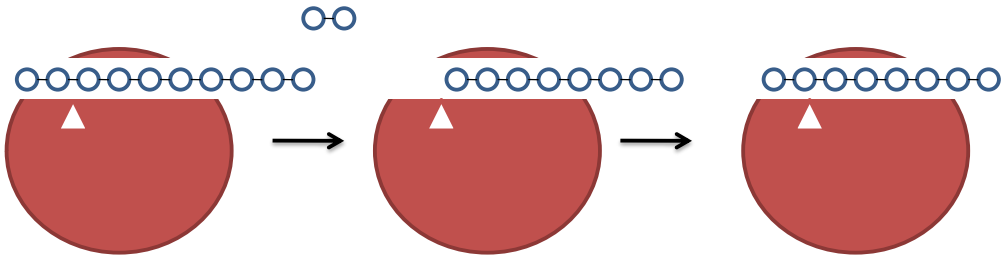


Fig. 2.2(ii). The mechanism of processivity of cellobiohydrolases.

2.5. Catalytic Mechanisms

2.5.1. Retaining and inverting glycosidases

The enzymatic catalysis of glycosidases proceeds by two distinct mechanistic classes based upon the stereochemical outcome. Inverting glycosidases, those hydrolyzing the glycosidic bond with net inversion of the anomeric configuration, and retaining glycosidase, those performing the catalysis with net retention (*9*) (Fig.2.3).

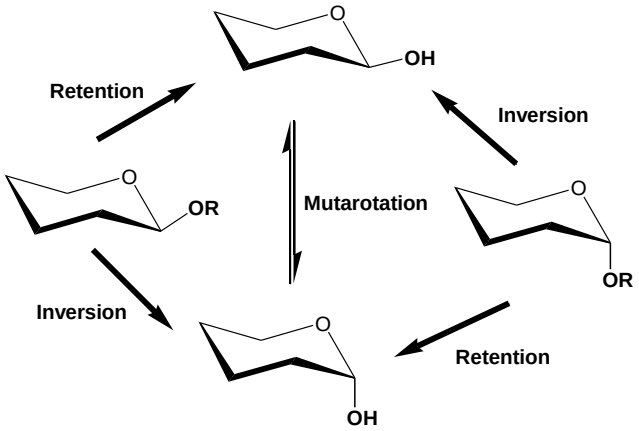


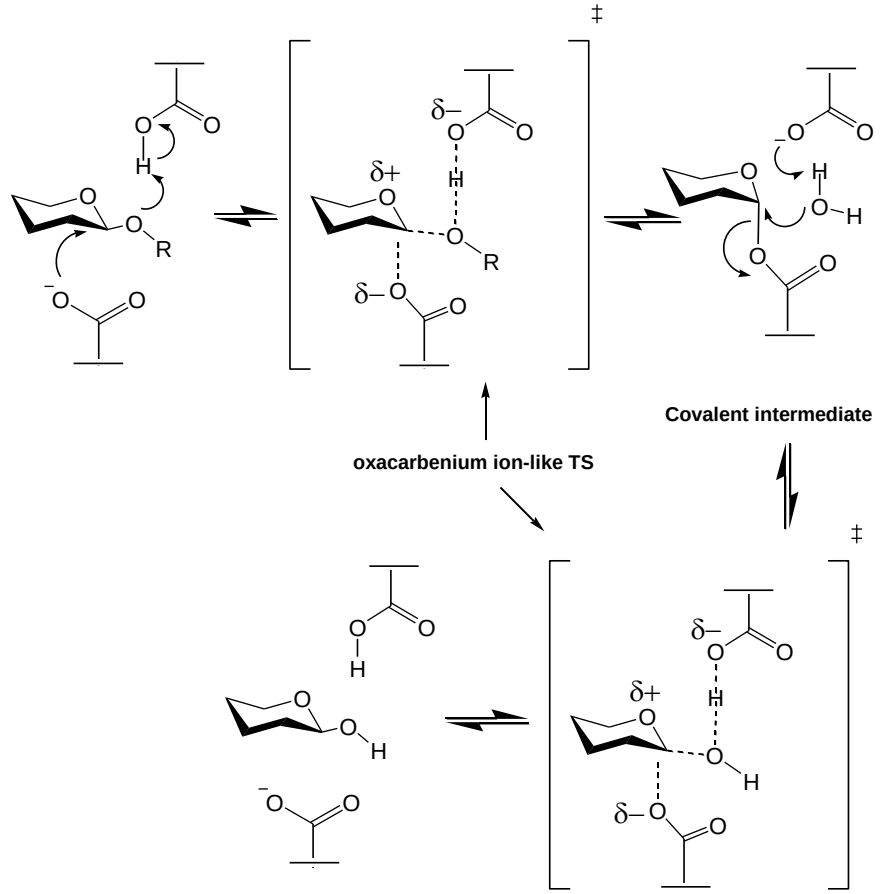
Fig.2.3. Glycosidases act with two possible stereochemical outcomes, net retention or inversion of the anomeric configuration. The mutarotation happens spontaneously.

The two mechanisms differ in that inverting glycosidases operate via a direct displacement of the leaving group by water, whereas retaining glycosidases utilize a double-displacement mechanism involving a glycosyl-enzyme intermediate (Fig. 2.4). In the retaining mechanism, the first step, cleavage of the glycosidic linkage, is called the glycosylation step, while the second step, attack of water, is referred to as deglycosylation.

In both the retaining and the inverting mechanisms, the position of the proton donor is identical, in other words it is within hydrogen-bonding distance of the glycosidic oxygen. In retaining enzymes, the nucleophile/base residue is in close vicinity of the sugar anomeric oxygen. This residue, however, is more distant in the inverting enzymes which must accommodate a water molecule between the sugar and the base residue. This difference explains the average distance between

the two catalytic residues of ~5.5 Å in the retaining enzymes against ~10 Å in the inverting enzymes (*10*).

a



b)

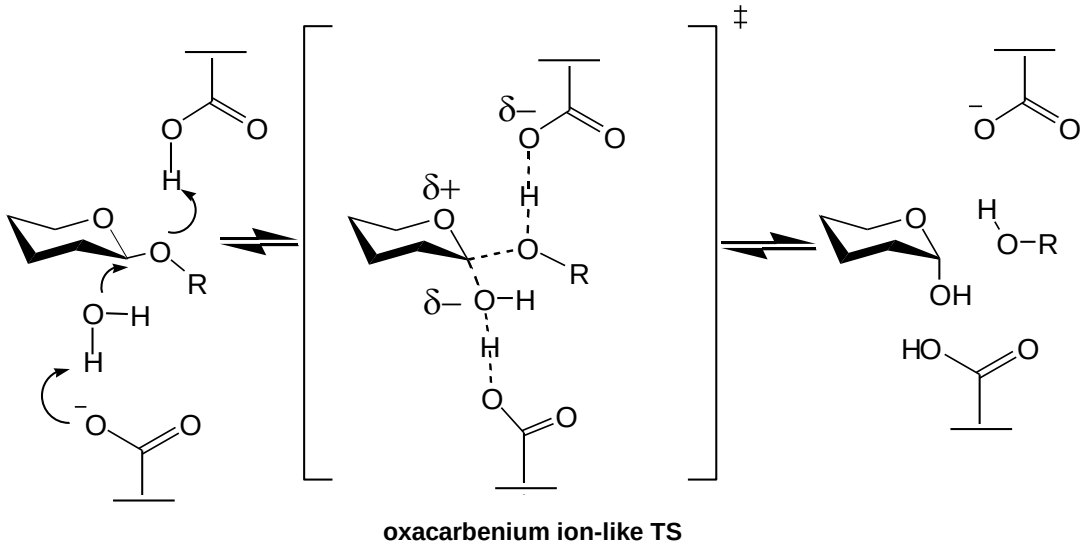


Fig.2.4 General mechanisms for retaining (a) and inverting (b) glycosidases.

2.5.2. Catalytic Carboxylates

Despite the differences between both classes, it is remarkable that both classes employ a pair of carboxylic acids at the active site. With inverting enzymes, one residue acts as a general acid and the other as a general base, whereas with retaining enzymes one functions as general acid/base while the other acts as a nucleophile/leaving group. In addition, both categories of enzymes operate via oxacarbenium ion-like transition states (Fig.2.4).

Interest in the structural features and mechanisms of these enzymes has increased in the past few years due to the potential use of glycosidases in diverse fields and applications, for instance, conversion of saccharidic biomass (primarily cellulose and starch), and in the design of glycosidase inhibitors in therapeutic applications for the treatment of cancer, diabetes and obesity, and of AIDS and other viral infections (*11*).

2.5.2.1. Alternative catalytic nucleophiles

It has been shown that the general catalytic acid/base residue is almost always an enzyme derived carboxylate. The nature of the nucleophile, however, has been found to vary in certain cases although most commonly it too is a carboxylate residue (Fig.2.5a). Notable exceptions to this topology are those enzymes from clan GH-E, which consists of GH families 33, 34, and 83 sialidases and trans-sialidases. These enzymes all perform catalysis using a retaining catalytic mechanism, and all have a tyrosine residue lying in close proximity to the anomeric carbon, where the carboxylate of most retaining glycosidases would be found (Fig.2.5b). Mass spectrometry studies of the *Trypanosoma cruzi* trans-sialidase (*12*) and detailed kinetic investigation of site-directed mutants of the *Micromonospora viridifaciens* sialidases (*13*) provided good support for this tyrosine acting as a catalytic nucleophile; observations that are consistent with the proposal made by Horenstein and coworkers on the basis of primary KIE studies (*14*). Interestingly, the structures have implicated a conserved enzymic carboxylate adjacent to the tyrosine as a likely catalytic base, which would act to enhance the nucleophilicity of the tyrosine residue. Little, however, is known about the consequences of deleting this carboxyl group although the effects should be pronounced. Validations for the evolution of such

an unusual ketal intermediate formed on a tyrosine residue within sialidases and trans-sialidases have been forwarded. Some explanations have been proposed to provide more understanding of the high intrinsic reactivity of sialosides (the sialosyl cation is considerably more stable than the glucopyranosyl cation). First, the sialosyl ester formed on a carboxylate is simply too reactive to enable the leaving group to diffuse away. The tyrosine nucleophile, which is a much worse leaving group, is thought to form a more stable tyrosine sialoside intermediate (*13*). The second proposal is that a tyrosine nucleophile avoids electrostatic repulsion of the anionic sialoside substrate that would occur if the enzymic nucleophile were a carboxylate (*12*). A third possible rationale is that the energetically important hydrogen bonding interaction of the nucleophile with the 2- hydroxyl group of hexopyranoside substrates (*15-17*) cannot occur with sialosides.

Another group of GHs that uses a nucleophile other than a carboxylate comprises various enzymes that act to cleave substrates having an acetamido group at the position adjacent to the reactive anomeric center. In this neighbouring-group (also termed substrate-assisted) catalytic mechanism the acetamido group is oriented and polarized by an enzymic residue to attack the anomeric center to form a covalent link between the anomeric center and the carbonyl oxygen of the acetamido group (Fig.2.5c). The number of families using this catalytic mechanism has increased from families 18, 20, and 56, to now include enzymes from family 84 (*18*).

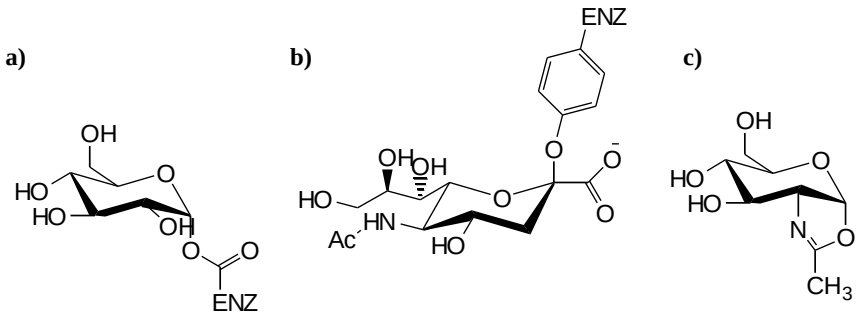


Fig.2.5. Some different glycosyl enzyme intermediates. (a) Covalent glycosyl enzyme intermediate formed on an enzymic carboxylate. (b) Covalent glycosyl enzyme intermediate formed on an enzymic tyrosine as observed for several trans-sialidases. (c) Oxazoline intermediate found for enzymes using substrate-assisted catalytis from the 2-acetamido group, the intermediate may also be an oxazolinium ion intermediate.

2.5.2.2. Unconventional catalytic pathways

A variation of the general mechanism for glycosidase enzymes has been demonstrated for different glycosidase members. Two glycoside hydrolases have recently been proposed to use unconventional mechanisms. A *Bacillus sp*. GH105, that acts on the terminal 4,5-ene product formed by the action of lyases, has been proposed to hydrate the alkene to generate a hemiketal center (Fig.2.6) (*19*).

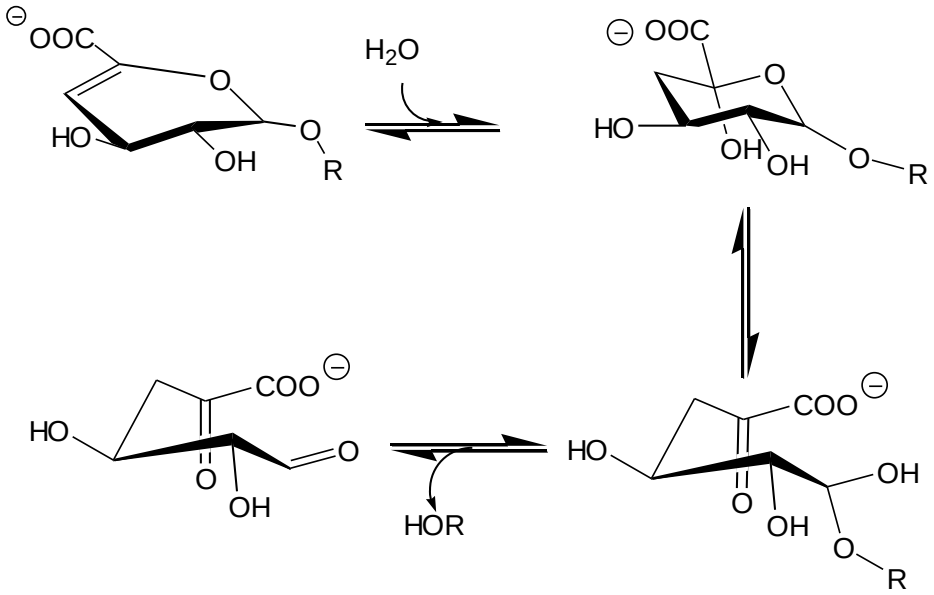


Fig.2.6. Unusual mechanism of Enzyme catalyzed hydration generates a hemiketal center. Ring opening generates the hemiacetal which spontaneously loses the aglycon to generate 4-deoxy-L-threo-5-hexosulose- uronate.

Another example of unusual mechanism was uncovered for the retaining GHs from families 4 and, recently, 109. Members of GH4 have long been known to be metal dependent and to require an NAD+ cofactor for activity, as well as unprecedented substrate promiscuity in that they can process both α-configured and β-configured substrates having activated leaving groups (*20*).

A series of structural and kinetic studies (*21, 22*) on the *Bacillus subtilis* (*23*) and two *Thermatoga maritime* (*24, 25*) enzymes revealed that they operate via oxidation and elimination mechanism to generate an enzyme sequestered enone intermediate that is hydrated and reduced to yield the product hemiacetal. Overall, this mechanistic treat reveals an unexpected motion of protons and electrons within the active site of a glycoside hydrolase.

While some members of family GH31 catalyze hydrolysis of axially linked pyranosides via the common retaining mechanism, some members proceed via unusual catalytic mechanism that yields 1,5-anhydro- D-fructose as the product of the reaction (Fig.2.7) (*26*).

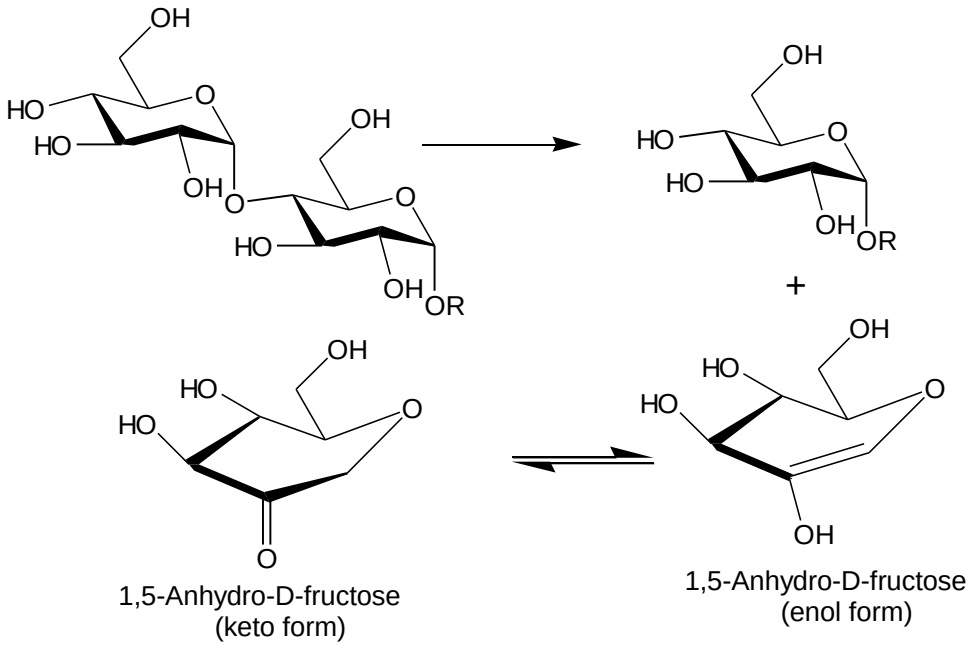


Fig.2.7. Unconventional mechanism proceeded by some members of family GH31 glycosidase

2.5.3. Rate-determining steps

To get understandings about the mechanistic pathways, it is critical first to identify a number of substrates for which different steps are rate-limiting. A number of aryl β-glycosidases have been successfully used for this purpose. Some studies have been performed on the almond β-glucosidase (*27*) and Escherichia coli β-galactosidase (*28*). In the case of *Agrobacterium* sp. β-glucosidase, a plot of log k_{cat} values for the hydrolysis of a series of aryl glucoside substrates versus the aglycone pK_a values produces a biphasic, concave-downward Brønsted relationship, indicating a change in rate-determining step as the aglycone leaving group ability increases (*29*) (Fig.2.5).

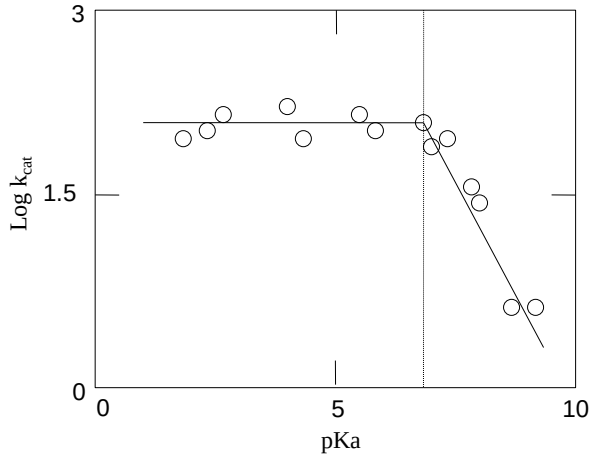


Fig. 2.8 Brønsted relationship for a series of aryl β-glucosides with *Agrobacterium* sp. β-glucosidase (Abg).

The glycosylation step is rate-limiting for substrates to the right of the break, as indicated by the strong dependence ($\beta_{lg} = -0.7$) in this region. Below the break, the k_{cat} value is independent of the leaving group, indicating that some other step is rate-limiting. This was shown to be the deglycosylation step through nucleophilic competition studies and the observation of "burst" kinetics in the pre-steady state (*17, 29*).

2.5.4. Noncovalent interactions

Noncovalent enzyme/substrate interactions that are optimized at the transition state could be a crucial component of catalysis in most enzymes. Hydrogen bond interactions at each sugar hydroxyl position with the enzyme active site has been probed for some β-glucosidases (*17*). In such studies, a series of deoxygenated and deoxyfluorinated derivatives of the substrate 2,4-dinitrophenyl β-glucoside were synthesized and subjected to detailed kinetic analysis. Key conclusions from this study were that binding interactions at the 3-, 4-, and 6-positions individually contribute 3−10 kJ mol-1 to each transition state, whereas contributions in the ground state are much weaker (≤3 kJ mol-1). However, the most interesting interactions are those at the 2-position, which contribute substantially (18−22 kJ mol-1) to stabilization of the glycosylation and deglycosylation transition states. This appears to be a common phenomenon with β-glycosidases with interactions at 2-position in some enzymes reaching 45 kJ mol^{-1}(*17*).

2.5.5. Transition state analysis

Both inverting and retaining enzymes catalyze glycoside hydrolysis via transition state with substantial oxacarbenium ion character (Fig.2.4). This oxacarbenium ion character has been shown to be present at the anomeric center in the transition states of the enzyme-catalyzed process, as also the case for the spontaneous hydrolysis mechanism. A direct means of measuring the extent of the positive charge development on the sugar has been established, involving measurement of the effects of electron-withdrawing substituents at the anomeric centre on the rates of enzymatic catalysis (*30*). This study aimed to trace changes in the transition state structure during evolution of the enzyme activity. Another study has been done to probe the degree of charge stabilization by the active site nucleophile by adopting kinetic analysis of mutants of *Agrobacterium* sp. β-glucosidase modified at this position (*31*).

Transition state structures have also been probed through the more conventional measurements of kinetic isotope effects. Some examples include studies of *Agrobacterium* sp. β-glucosidase(*29*) and *C.fimi* exoglycanase (*32*). Combination of isotope effect measurements with computational methods in a vibrational analysis has provided deeper insights into transition state structures. This has been applied to nucleoside hydrolases (*33, 34*). Another fascinating series of such studies have provided direct insights into the evolution of enzymatic activity, in which the changes in the transition state structure of a vestigial *E. coli* β-galactosidase have been probed as it evolved under selective pressure (*30, 35*).

Another important aspect of transition state analysis is the measurement of the binding interaction present between the individual sugar hydroxyl groups and the enzyme at the transition state. These have been assessed in several cases through kinetic analysis of specifically deoxygenated substrates. Values ranging from 10 to 18 kJ mol^{-1} have been measured for the 4 and 6 hydroxyls of maltose in glucoamylase (*36*), and from about 20 kJ mol^{-1} for the 3, 4- and 6-hydroxyls to up to 33 kJ mol^{-1} for the 2-hydroxyl of galactose in β-galactosidase (*37*). These binding interactions can be quite remote from the bond being cleaved. For instance, binding interactions with the distal sugar moiety in cellobiosides preferentially stabilize the glycosylation step of *C.fimi* β-1, 4-glycanase (*32*).

2.5.6. Covalent intermediate

The nature of the intermediate as a covalent glycosyl enzyme species has obtained extensive support, first in a classical paper by Sinnott and Souchard using kinetic isotope effects (*28*), and since then by using a wide variety of methods (*38-42*). The best evidence for the covalent glycosyl enzyme intermediate stems from studies making use of kinetic isotope effects. The normal and generally large $^{\alpha\text{-D}}$(V) KIEs measured for both the glycosylation and deglycosylation steps can only be easily reconciled with the formation of a transient glycosyl enzyme intermediate that is flanked by oxacarbenium ion-like transition states in which the anomeric center is trigonal, thus having greater sp^2-hybridization than the intermediate.

A clearer picture on the nature of the covalent glycosyl enzyme intermediate has been shed by adopting the 2-fluorosugar strategy, in which a glycosyl-enzyme intermediate is trapped through treatment with an activated 2-deoxy-2-fluoroglycoside (the fluorine serving to destabilise the oxacarbenium ion transition states). It proved possible to accumulate different deoxy-fluoroglycosyl enzyme intermediates whose lifetimes were prolonged sufficiently to allow crystallographic studies (*43-45*). Since such substituents slow both glycosylation and deglycosylation steps, a good leaving group, typically fluoride or dinitrophenolate, is generally incorporated into the analogue to ensure that the glycosylation step is faster than the deglycosylation, and thus the intermediate accumulates.

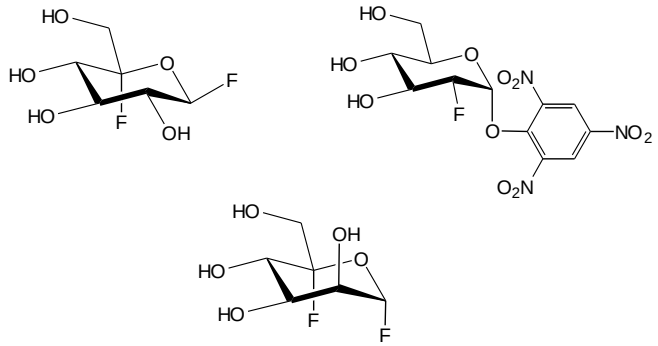


Fig.2.9. Examples of fluorosugars used for trapping the glycosyl enzyme intermediate.

Certainly the best insight into the nature of the glycosyl enzyme intermediate has come from X-ray crystallographic analysis of trapped complexes. Structures of

several such complexes have been now been solved (*46-49*), the first being that of the 2-deoxy-2-fluorocellobiosyl enzyme intermediate of the retaining β-1,4-glycanase Cex from Cellulomonas fimi (*50*).

2.6. Substrate ring distortion

The role of substrate distortion in the mechanism among glycosidases is a contentious issue with a long history. Enzyme-conferred distortion was first proposed by David Phillips, during his seminal studies on hen-egg white lysozyme, albeit in the context of S_N1-like mechanisms via a stable cationic intermediate.

For the enzyme catalytic activity to proceed, transition state structures should have a planar geometry in which the C-5, O-5, C-1, and C-2 atoms adopt a coplanar arrangement. This coplanarity permits efficient orbital overlap and delocalization of electrons between the lone pair of the endocyclic ring oxygen with the transient orbital of the anomeric centre (Fig. 2.10a). It was recognized that such planarity would limit the number of available conformations the pyranose ring could adopt and Sinnott provided evidence for this hypothesis using a series of secondary α-deuterium and β-deuterium KIE experiments (*9*).

For many years, distortion invariably was interpreted as referring to ‘sofa’ (more formally ‘envelope’) or half-chair conformations. These transition states, however, can actually adopt one of two half-chair ($^4H_3/^3H_4$ or their closely related ^4E and ^3E envelopes) or boat ($^{2,5}B/B_{2,5}$) conformations (Fig.2.10b).

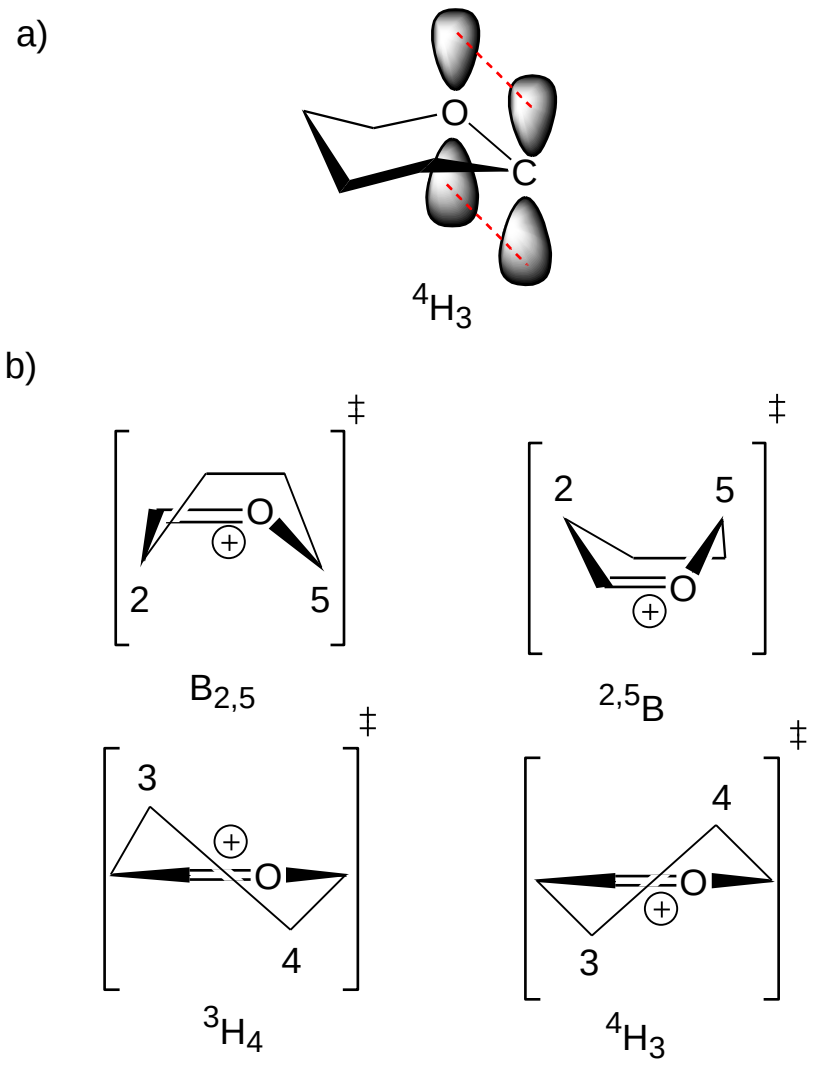


Fig.2.10. a) Stabilization of the developing positive charge at the anomeric center by the p-like lone pair of the endocylic oxygen favors the oxacarbenium ion-like transition state adopting certain conformations. b) Possible conformations of oxacarbenium ion-like transition states.

The first structurally analyzed examples of distorted 'Michaelis' complexes (so named after the formation of the Michaelis complex in classical enzyme kinetics) are two structures of β-D-glycosidases revealing intact substrates distorted away from their favoured $^4C_1$ chair conformations toward envelope and skew-boat conformations (*51, 52*). The prominent feature of both of these complexes is the distortion of substrate such that the leaving group is positioned in an axial orientation that allows in-line nucleophilic attack at the σ* anti-bonding orbital of the anomeric carbon. Moreover, the conformations of the substrates adopted in these complexes both reduce steric hindrance to nucleophilic attack by H-1 and

prevent 1,3-diaxial clashes of the nucleophile with H-3 and H-5. These conformations, and many observed since (*46, 53*), are also consistent with participation of lone pair electrons on the endocyclic oxygen. Furthermore, these structures gave hints that one could predict transition state structures on the basis of careful observations of enzyme complexes in which the conformations of stable species along the reaction coordinate were well defined.

The first distorted covalent intermediate was reported for GH31 α-glucosidase (*54*). In this study, the intermediate adopted the $^{1}S_{3}$ conformation. Similarly, trapped intermediates of hydrolytic α-mannosidases (*55*) and α-L-fucosidases (*56*) have also revealed distorted intermediates consistent with the need for in-line attack of nucleophilic water.

More recently, there have been reports of a number of trapped covalent tyrosyl-enzyme intermediates for sialidases — all show $^{2}C_{5}$ conformations (equating to $^{1}C_{4}$ for aldohexopyranosides) for the intermediate (*57*) including a notable atomic resolution analysis by the Taylor group (*58*). All of these structures, however, indicate catalysis very likely proceeds via the same $^{4}H_{5}$ transition state conformation (equating to $^{3}H_{4}$ for an aldohexopyranose).

The conformational pathway, in other words, pseudorotational conformational itinerary for pyranose sugar ring has been proposed a while ago by Stoddart (*59*). He established a diagram which shows schematically all the accessible conformers for a single α-glucopyranose unit, according to their corresponding IUPAC nomenclature.

The itinerary map of Stoddart still gives the possible conformational pathways that a pyranose ring may follow as it moves from one conformation to another. However, energetically no information about relative stability of different conformation can be extracted, nor can it be assumed that all conformations on this map correspond to stationary points on the free energy landscape with respect to ring distortion. Hence, A detailed description of the free energy pathways between all the conformations of Stoddart's diagram would be useful. Many studies of glycosidase mechanisms have represented the relative energy of sugar ring conformations in terms of Stoddart diagram (*53, 60, 61*) (Fig.2.11).

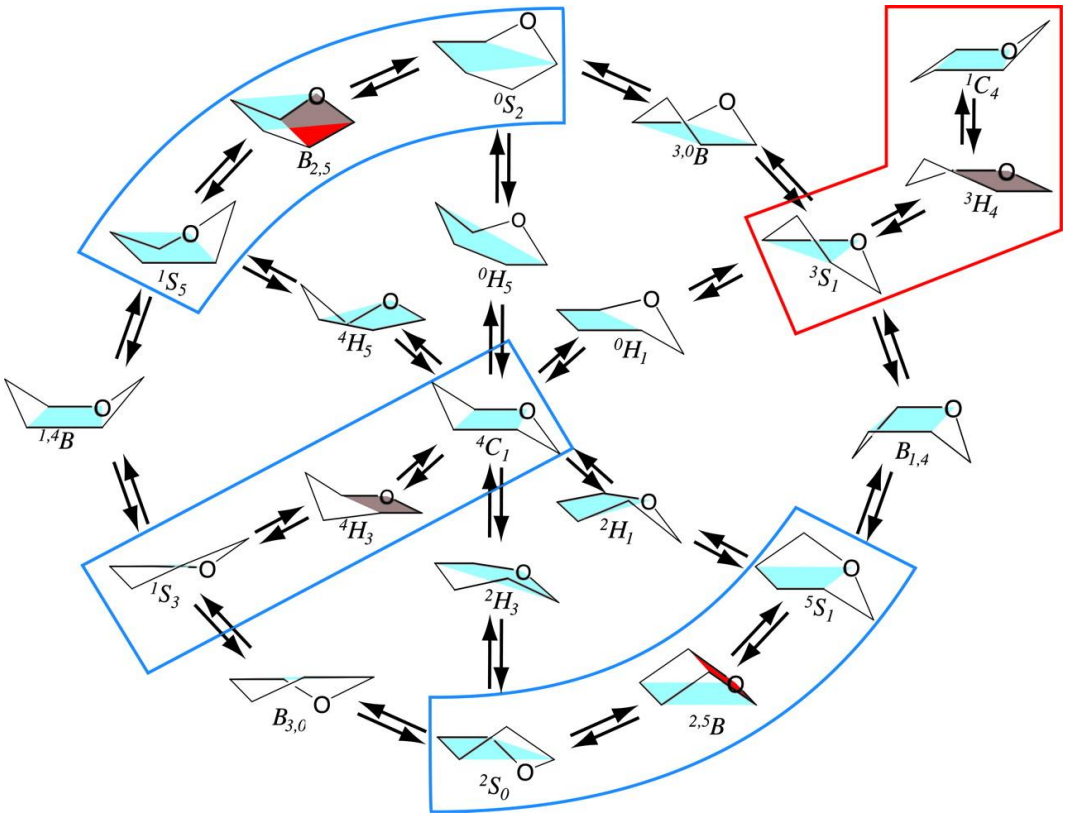


Fig.2.11. On the light of Stoddart's map, Pseudorotational itinerary for the interconversion of sugar ring conformations highlighting the proposed transition states employed by glycosidases.

Different theoretical approaches have been adopted to quantify the energy differences among glucopyranose ring conformations (*62-67*) Most of these calculations are based on the use of force fields and thus electronic rearrangements cannot be analyzed. In recent years, density functional theory (DFT) has shown itself to be a powerful technique to address this and related issues (*68-70*).

In the same regard, a recent unique study presents a free energy landscape of the skeleton of a glucopyranose unit, corresponding to Stoddart's diagram (*71*). This approach takes into account (except for sampling errors) the full configurational contribution, including conformational transitions of the side groups. The calculations adopted in this investigation are based upon Car-Parrinello molecular dynamics (CPMD), combined with the recently developed

metadynamics approach, which is aimed at enhancing the sampling of the phase space and at mapping the free energy landscape as a function of a small number of collective variables. Collective variables represented as the puckering coordinates, which introduced by Cremer and Pople (*72*), to construct the energy map of Stoddart's diagram for α-D-glucopyranose. Analysis of the structural and electronic rearrangements along the conformational itinerary has been also probed in this study.

Finally, it is worth mentioning here that, among the potential pseudorotational conformational itineraries for sugar pyranose ring interconversions, there are four potential conformations where the planarity of C-5, O-5, C-1, and C-2 is satisfied (Fig.2.11, $^{2,5}B$, $B_{2,5}$, $^{4}H_{3}$, and $^{3}H_{4}$) (*60*). Recent data have suggested that three of these potential transition state conformations ($^{2,5}B$, $B_{2,5}$, and $^{4}H_{3}$) are employed among the known GH families (*60,73-75*). Evidence favouring these conformations has come from the analysis of x-ray structures of target enzymes with trapped covalent intermediates using fluorinated substrate analogs (for retaining glycosidases), with enzyme bound nonhydrolyzable substrate mimics, or with natural substrates in combination with active site mutants (*60, 73-75*). Of the four possible coplanar transition states, only the $^{3}H_{4}$ conformation has not yet been identified as an intermediate in any glycosidase mechanism.

However, all the previous studies have provided insights for understanding this distinctive feature among glycosidases, none of them has explored how the enzyme could stabilize the distorted sugar conformation and the energetic cost that the enzyme needs to afford. In this work, via a unique scenario, we took one step forward in that direction to answer these questions and shed a new spot on this area.

2.7. α versus β- glycosidases

Kinetic studies from the trapping of glycosyl intermediates (*76*) and from the three-dimensional structure of such intermediates (*40*) have provided a several lines of evidence point toward a subtle difference in the oxacarbenium ion character of the transition structures formed on α- and β-glycosiadases. The difference is based on the interaction pattern between the nucleophilic carboxylate, anomeric centre and the 2-hydroxyl group. On β-glycosidases, the

syn interaction of the nucleophilic carboxyl oxygens with the anomeric centre and the 2-hydroxyl will tend to favour a greater share of the positive charge on the anomeric carbon (Fig.2.12a). This interaction is not possible in case of α-glycosidases, where instead a *syn* interaction of the nucleophile carboxyl oxygens with the anomeric and the endocyclic oxygen is observed in the intermediate. This interaction will favour positive charge development on the endocyclic oxygen (Fig.2.12b)

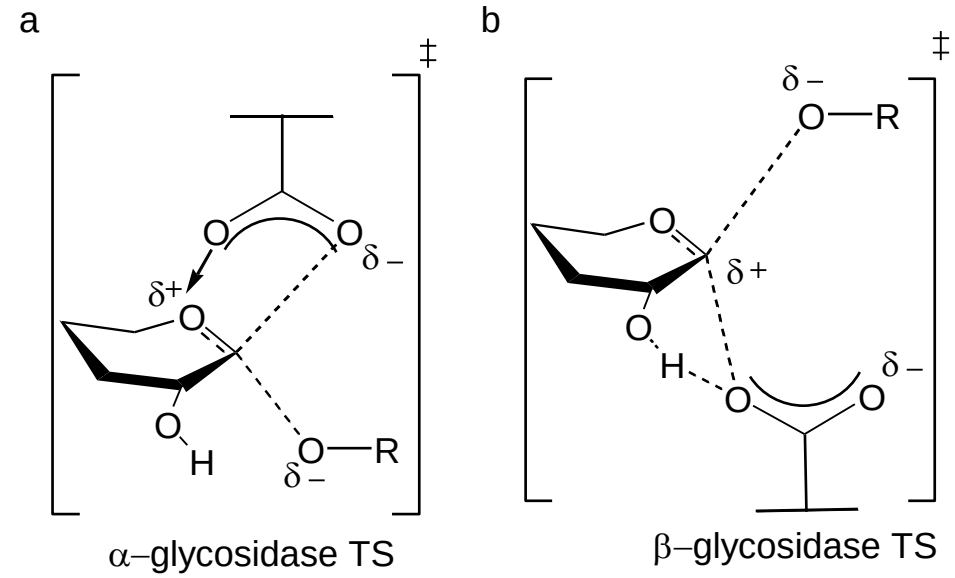


Fig.2.12. Comparison of the transition states for α– (a) and β- (b) retaining glycosidases.

This hypothesis has been supported by some experimental observations using some inhibitors, summarized as follows. The efficacy of the 2-fluorosugar inhibitors with β-glycosidases, but not with α-glycosidases, could be due to the fact that only with β-glycosidases, wherein the charge is more localized to a greater extent on the anomeric centre, is there sufficient inductive destabilization to suppress the rate significantly. By contrast, 5-fluoro inhibitors function with both α- and β-glycosidases.

2.8. Motifs towards designing of glycosidases inhibitors

Glycoside hydrolases are some of the most proficient of enzymes, accelerating the spontaneous reactions they catalyze by up to 10^{17}-fold (*77*). This suggests

that very tight binding transition state analogue inhibitors of glycosidases are possible, with potential as therapeutic agents (*78*) . In the following text, I highlight some of these therapeutic targets.

Much interest has been paid towards glycosidase inhibitors as therapeutics in recent years with the successful marketing of the α-amylase inhibitor acarbose for control of blood glucose levels and of the sialidase inhibitor Relenza and Tamiflu as a treatment for influenza.

Recently, there have been greater efforts to understand the fundamental basis of inhibition of GHs by a number of effective inhibitors (*79, 80*). For many years, synthesizing 'transition state analogues' have been the customary motif towards designing potent inhibitors. However, no good evidence exists to support such an assignment other than high, or sometimes even modest, inhibitory potency. Interestingly, a combined structural and molecular modelling study has suggested that the pursuit of the transition state analogy is not the only route to high potency. In this study, Golgi α-mannosidase inhibitor mannostatin (Fig.2.13a), which resembles not the transition state but rather the covalent glycosyl enzyme intermediate (Fig.2.13b), has a remarkable potency (K_I = 36 nM) (*81*).

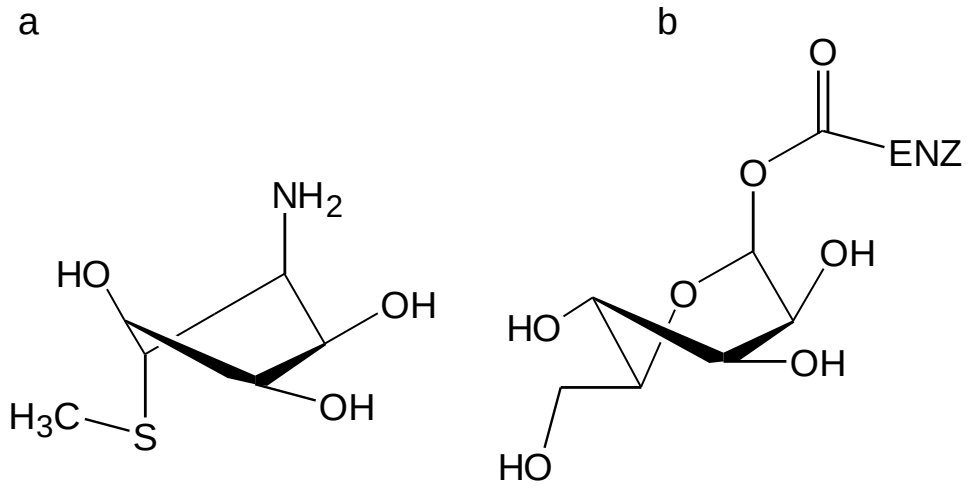


Fig.2.13. (a) Mannostatin adopts a $^{2}T_{1}$ conformation that resembles the $^{1}S_{5}$ conformation of the mannosyl enzyme intermediate of (b) Drosophila Golgi α-mannosidase.

In the same regard, the transition state analogy of aminosugars is also currently a topic of vagueness. The valienamine class of inhibitors, which includes Acarbose, are clearly effective inhibitors but their resemblance to the transition

state of GH13 amylases remains unclear. One group proposes Acarbose (Fig.2.14a) is a good transition state analog (*82*) while another has suggested it is a poor mimic of the transition state (*83*). The $^{2}H_{3}$ half- chair conformation of valienamine (Fig.2.14b), however, is certainly not one that could be adopted by an oxacarbenium ion-like transition state.

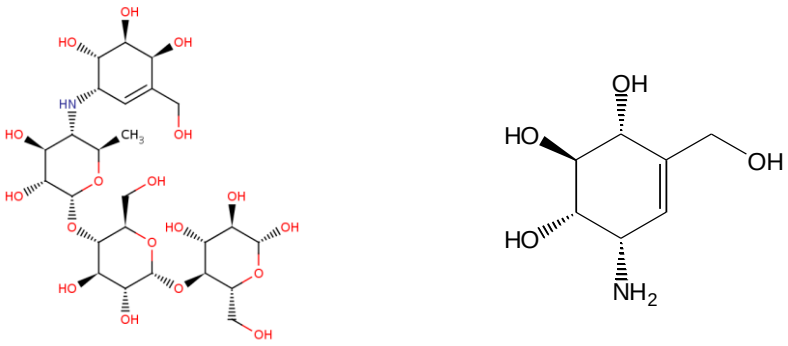


Fig.2.14. 2-D sketches of acarbose, (a), and valienamine, (b).

Furthermore, the inhibition of GH20 chitinases by cyclic peptides (*84*) and continuing isolation of potent GH13 amylase inhibitors (*85*) underscore that GHs may be inhibited by a wide variety of different structures.

Generally speaking, enzyme selectivity is a critical aspect that should be considered carefully during designing of inhibitors. In glycosidases side, it has been reported that small changes in inhibitor design can confer great selectivities that facilitate in vivo studies (*86*). Emergence of the recent structure of the H5N1 Tamiflu resistant sialidase (*87*) should provoke the need of new classes of effective inhibitors to target enzymes of medical interest.

Further refinements to our understanding of the structures of transition states and the thermodynamic subtleties of highly polar molecules binding to GHs should facilitate the generation of more potent and selective inhibitors.

Despite the huge effort that has been done to explore the catalytic mechanisms of glycoside hydrolases, this area still an exciting field to explore and deeper mechanistic insights will probably require greater collaborative efforts. New strategies will be required (*88*) to carry out detailed and comprehensive kinetic isotope effects, analogous to those carried out on N-ribosyltransferases, which have enabled modeling of transition state structures and the development of extremely potent inhibitors (for review see Schramm (*89*)).

2.9. Theoretical studies on GHs

Theoretical studies on the structures and dynamics of GH enzymes have been increasing during the past decade. Ring distortion, catalytic mechanisms, characterization of transition states have attracted much interest because inhibitors of these enzymes can act as new therapeutic agents for treatment of many diseases particularly viral infections (*90, 91*).

Substrate sugar ring distortion has been the subject of many recent theoretical studies. Classical molecular dynamics simulations demonstrated that the boat conformation at subsite -1 is critical in the mechanism of family 18 chitinases (*92*). Recent studies confirmed that the -1 sugar moiety of the substrate in cellulose Ce16A from *Trichoderma reesei*, an inverting glycosidase, adopts a skew-boat conformation ($^2S_0$) (*93*). Similarly, modelling studies of β-galactosidases provided evidence of substrate distortion (*94*). All these results rely on parameterized force fields to describe the interaction among atoms, and thus the interplay of electronic/structural factors on the substrate conformation cannot be analyzed. A recent study has taken into account the electronic effects and charge rearrangements in the active site by means of hybrid QM/MM simulation to investigate the conformation of the substrate in the Michaelis complex of *Bacillus* 1,3-1,4-β-glucanase (*95*). Recent studies demonstrated the conformational itinerary and relative energy of ring conformations of glucopyranose (*71*).

Understanding the GH catalytic mechanisms and characterizing transition states has been a target of many computational studies. A theoretical rational of the formation of covalent-enzyme intermediate among glycosidases has been probed using different theoretical approaches (*96-102*). Investigation into the structure and role of oxacarbenium ion in GH mechanisms provides a deeper insight into the catalytic pathway of GH enzymes (*103-104*). Knowledge of the reaction mechanism and transition states involved in the catalytic route is important for the design of transition state analogs, which may act as potential inhibitors. Theoretical modelling of transition state analogs proved a successful approach towards designing of potential antiviral drugs (*91, 92*).

References

1. Henrissat, B. (1991) A Classification of Glycosyl Hydrolases Based on Amino-Acid-Sequence Similarities, *Biochemical Journal 280*, 309-316.
2. Henrissat, B., and Bairoch, A. (1993) New Families in the Classification of Glycosyl Hydrolases Based on Amino-Acid-Sequence Similarities, *Biochemical Journal 293*, 781-788.
3. Torronen, A., Kubicek, C. P., and Henrissat, B. (1993) Amino-Acid-Sequence Similarities between Low-Molecular-Weight Endo-1,4-Beta-Xylanases and Family-H Cellulases Revealed by Clustering Analysis, *Febs Letters 321*, 135-139.
4. Divne, C., Stahlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T., and Jones, T. A. (1994) The 3-Dimensional Crystal-Structure of the Catalytic Core of Cellobiohydrolase-I from Trichoderma-Reesei, *Science 265*, 524-528.
5. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1995) Conserved Catalytic Machinery and the Prediction of a Common Fold for Several Families of Glycosyl Hydrolases, *Proceedings of the National Academy of Sciences of the United States of America 92*, 7090-7094.
6. Jenkins, J., Leggio, L. L., Harris, G., and Pickersgill, R. (1995) Beta-Glucosidase, Beta-Galactosidase, Family-a Cellulases, Family-F Xylanases and 2 Barley Glycanases Form a Superfamily of Enzymes with 8-Fold Beta/Alpha-Architecture and with 2 Conserved Glutamates near the Carboxy-Terminal Ends of Beta-Strand-4 and Beta-Strand-7, *Febs Letters 362*, 281-285.
7. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases, *Biochemical Journal 321*, 557-559.
8. Davies, G., and Henrissat, B. (1995) Structures and Mechanisms of Glycosyl Hydrolases, *Structure 3*, 853-859.
9. Sinnott, M. L. (1990) Catalytic Mechanisms of Enzymatic Glycosyl Transfer, *Chemical Reviews 90*, 1171-1202.

10. Mccarter, J. D., and Withers, S. G. (1994) Mechanisms of Enzymatic Glycoside Hydrolysis, *Current Opinion in Structural Biology 4*, 885-892.
11. Hughes, A. B., and Rudge, A. J. (1994) Deoxynojirimycin - Synthesis and Biological-Activity, *Natural Product Reports 11*, 135-162.
12. Watts, A. G., Damager, I., Amaya, M. L., Buschiazzo, A., Alzari, P., Frasch, A. C., and Withers, S. G. (2003) Trypanosoma cruzi trans-sialidase operates through a covalent sialyl-enzyme intermediate: Tyrosine is the catalytic nucleophile, *Journal of the American Chemical Society 125*, 7532-7533.
13. Watson, J. N., Dookhun, V., Borgford, T. J., and Bennet, A. J. (2003) Mutagenesis of the conserved active-site tyrosine changes a retaining sialidase into an inverting sialidase, *Biochemistry 42*, 12682-12690.
14. Yang, J. S., Schenkman, S., and Horenstein, B. A. (2000) Primary C-13 and beta-secondary H-2 KIEs for trans-sialidase. A snapshot of nucleophilic participation during catalysis, *Biochemistry 39*, 5902-5910.
15. Zechel, D. L., and Withers, S. G. (2000) Glycosidase mechanisms: Anatomy of a finely tuned catalyst, *Accounts of Chemical Research 33*, 11-18.
16. Roeser, K. R., and Legler, G. (1981) Role of Sugar Hydroxyl-Groups in Glycoside Hydrolysis - Cleavage Mechanism of Deoxyglucosides and Related Substrates by Beta-Glucosidase-A3 from Aspergillus-Wentii, *Biochimica Et Biophysica Acta 657*, 321-333.
17. Namchuk, M. N., and Withers, S. G. (1995) Mechanism of Agrobacterium .beta.-glucosidase: kinetic analysis of the role of noncovalent enzyme/substrate interactions, *Biochemistry 34*, 16194-16202.
18. Macauley, M. S., Whitworth, G. E., Debowski, A. W., Chin, D., and Vocadlo, D. J. (2005) O-GlcNAcase uses substrate-assisted catalysis - Kinetic analysis and development of highly selective mechanism-inspired inhibitors, *Journal of Biological Chemistry 280*, 25313-25322.
19. Itoh, T., Hashimoto, W., Mikami, B., and Murata, K. (2006) Crystal structure of unsaturated glucuronyl hydrolase complexed with substrate, *Journal of Biological Chemistry 281*, 29807-29816.

20. Thompson, J., Pikis, A., Ruvinov, S. B., Henrissat, B., Yamamoto, H., and Sekiguchi, J. (1998) The gene glvA of Bacillus subtilis 168 encodes a metal-requiring, NAD (H)-dependent 6-phospho-alpha-glucosidase - Assignment to family 4 of the glycosylhydrolase superfamily, *Journal of Biological Chemistry 273*, 27347-27356.

21. Rajan, S. S., Yang, X. J., Collart, F., Yip, V. L. Y., Withers, S. G., Varrot, A., Thompson, J., Davies, G. J., and Anderson, W. F. (2004) Novel catalytic mechanism of glycoside hydrolysis based on the structure of an NAD(+)/Mn2+-dependent phospho-alpha-glucosidase from Bacillus subtilis, *Structure 12*, 1619-1629.

22. Varrot, A., Yip, V. L. Y., Li, Y. S., Rajan, S. S., Yang, X. J., Anderson, W. F., Thompson, J., Withers, S. G., and Davies, G. J. (2005) NAD(+) and metal-ion dependent hydrolysis by family 4 glycosidases: Structural insight into specificity for phospho-beta-D-glucosides, *Journal of Molecular Biology 346*, 423-435.

23. Yip, V. L. Y., Thompson, J., and Withers, S. G. (2007) Mechanism of GlvA from Bacillus subtilis: A detailed kinetic analysis of a 6-phospho-alpha-glucosidase from glycoside hydrolase family 4, *Biochemistry 46*, 9840-9852.

24. Yip, V. L. Y., Varrot, A., Davies, G. J., Rajan, S. S., Yang, X. J., Thompson, J., Anderson, W. F., and Withers, S. G. (2004) An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 beta-glycosidase from Thermotoga maritima, *Journal of the American Chemical Society 126*, 8354-8355.

25. Yip, V. L. Y., and Withers, S. G. (2006) Mechanistic analysis of the unusual redox-elimination sequence employed by Thermotoga maritima BgIT: A 6-phospho-beta-glucosidase from glycoside hydrolase family 4, *Biochemistry 45*, 571-580.

26. Yu, S. K., and Marcussen, J. (1999) alpha-1,4-glucan lyase, molecular features and its use for production of 1,5-anhydro-D-fructose from starch, *Recent Advances in Carbohydrate Bioengineering*, 243-250

27. Nath, R. L., Rydon, H. N. The influence of structure on the hydrolysis of substituted phenyl β-D-glucosides by emulsin. *Biochem. J.* **1954**, *57*, 1−10.

28. Sinnott, M. L., Souchard, I. J. L. The mechanism of action of β-galactosidase. *Biochem. J.* **1973**, *133*, 89−98.

29. Kempton, J. B., and Withers, S. G. (1992) Mechanism of Agrobacterium .beta.-glucosidase: kinetic studies, *Biochemistry 31*, 9961-9969.

30. Srinivasan, K., Konstantinidis, A., Sinnott, M. L., and Hall, B. G. (1993) Large Changes of Transition-State Structure during Experimental Evolution of an Enzyme, *Biochemical Journal 291*, 15-17.

31. Withers, S. G., Rupitz, K., Trimbur, D., and Warren, R. A. J. (1992) Mechanistic Consequences of Mutation of the Active-Site Nucleophile Glu-358 in Agrobacterium Beta-Glucosidase, *Biochemistry 31*, 9979-9985.

32. Tull, D., and Withers, S. G. (1994) Mechanisms of Cellulases and Xylanases - a Detailed Kinetic-Study of the Exo-Beta-1,4-Glycanase from Cellulomonas-Fimi, *Biochemistry 33*, 6363-6370.

33. Horenstein, B. A., and Schramm, V. L. (1993) Correlation of the Molecular Electrostatic Potential Surface of an Enzymatic Transition-State with Novel Transition-State Inhibitors, *Biochemistry 32*, 9917-9925.

34. Horenstein, B. A., and Schramm, V. L. (1993) Electronic Nature of the Transition-State for Nucleoside Hydrolase - a Blueprint for Inhibitor Design, *Biochemistry 32*, 7089-7097.

35. Elliott, A. C., Srinivasan, K., Sinnott, M. L., Smith, P. J., Bommuswamy, J., Guo, Z., Hall, B. G., and Zhang, Y. L. (1992) The Catalytic Consequences of Experimental Evolution - Studies on the Subunit Structure of the 2nd (Ebg) Beta-Galactosidase of Escherichia-Coli, and on Catalysis by Ebgab, an Experimental Evolvant Containing 2 Amino-Acid Substitutions, *Biochemical Journal 282*, 155-164.

36. Sierks, M. R., Bock, K., Refn, S., and Svensson, B. (1992) Active-Site Similarities of Glucose-Dehydrogenase, Glucose-Oxidase, and Glucoamylase Probed by Deoxygenated Substrates, *Biochemistry 31*, 8972-8977.

37. Mccarter, J. D., Adam, M. J., and Withers, S. G. (1992) Binding-Energy and Catalysis - Fluorinated and Deoxygenated Glycosides as Mechanistic Probes of Escherichia-Coli (Lacz) Beta-Galactosidase, *Biochemical Journal 286*, 721-727.

38. Bause, E., and Legler, G. (1980) Isolation and structure of a tryptic glycopeptide from the active site of [beta]-glucosidase A3 from Aspergillus wentii, *Biochimica et Biophysica Acta (BBA) - Protein Structure 626*, 459-465.

39. Withers, S. G., Street, I. P., Bird, P., and Dolphin, D. H. (1987) 2-Deoxy-2-fluoroglucosides: a novel class of mechanism-based glucosidase inhibitors, *Journal of the American Chemical Society 109*, 7530-7531.

40. Uitdehaag, J. C. M., Mosi, R., Kalk, K. H., van der Veen, B. A., Dijkhuizen, L., Withers, S. G., and Dijkstra, B. W. (1999) X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the alpha-amylase family, *Nature Structural Biology 6*, 432-436.

41. Notenboom, V., Birsan, C., Nitz, M., Rose, D. R., Warren, R. A. J., and Withers, S. G. (1998) Insights into transition state stabilization of the beta-1,4-glycosidase Cex by covalent intermediate accumulation in active site mutants, *Nature Structural Biology 5*, 812-818.

42. Vocadlo, D. J., Davies, G. J., Laine, R., and Withers, S. G. (2001) Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate, *Nature 412*, 835-838.

43. McCarter, J. D., Yeung, W., Chow, J., Dolphin, D., and Withers, S. G. (1997) Design and synthesis of 2'-deoxy-2'-fluorodisaccharides as mechanism-based glycosidase inhibitors that exploit aglycon specificity, *Journal of the American Chemical Society 119*, 5792-5797.

44. Withers, S. G., McCarter, J. D., Howard, S., and Vocadlo, D. (1997) 5-fluoro-glycosyl fluorides as probes of glycosidases., *Abstracts of Papers of the American Chemical Society 214*, 4-Carb.

45. McCarter, J. D., and Withers, S. G. (1996) 5-fluoro glycosides: A new class of mechanism-based inhibitors of both alpha- and beta-glucosidases, *Journal of the American Chemical Society 118*, 241-242.

46. Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schulein, M., and Withers, S. G. (1998) Snapshots along an enzymatic reaction coordinate: Analysis of a retaining beta-glycoside hydrolase, *Biochemistry 37*, 11707-11713.

47. Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., and Henrissat, B. (1997) The crystal structures of Sinapis alba myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase, *Structure 5*, 663-675.

48. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase, *Biochemistry 38*, 5346-5354.

49. Notenboom, V., Birsan, C., Warren, R. A. J., Withers, S. G., and Rose, D. R. (1998) Exploring the cellulose/xylan specificity of the beta-1,4-glycanase Cex from Cellulomonas fimi through crystallography and mutation, *Biochemistry 37*, 4751-4758.

50. White, A., Tull, D., Johns, K., Withers, S. G., and Rose, D. R. (1996) Crystallographic observation of a covalent catalytic intermediate in a beta-glycosidase, *Nature Structural Biology 3*, 149-154.

51. Sulzenbacher, G., Driguez, H., Henrissat, B., Schulein, M., and Davies, G. J. (1996) Structure of the Fusarium oxysporum endoglucanase I with a nonhydrolyzable substrate analogue: Substrate distortion gives rise to the preferred axial orientation for the leaving group, *Biochemistry 35*, 15280-15287.

52. Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S., and Vorgias, C. E. (1996) Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease, *Nature Structural Biology 3*, 638-648.

53. Money, V. A., Smith, N. L., Scaffidi, A., Stick, R. V., Gilbert, H. J., and Davies, G. J. (2006) Substrate distortion by a lichenase highlights the different conformational itineraries harnessed by related glycoside hydrolases, *Angewandte Chemie-International Edition 45*, 5136-5140.

54. Lovering, A. L., Lee, S. S., Kim, Y. W., Withers, S. G., and Strynadka, N. C. J. (2005) Mechanistic and structural analysis of a family 31 alpha-glycosidase and its glycosyl-enzyme intermediate, *Journal of Biological Chemistry 280*, 2105-2115.

55. Numao, S., Kuntz, D. A., Withers, S. G., and Rose, D. R. (2003) Insights into the mechanism of Drosophila melanogaster Golgi alpha-mannosidase II through the structural analysis of covalent reaction intermediates, *Journal of Biological Chemistry 278*, 48074-48083.

56. Sulzenbacher, G., Bignon, C., Nishimura, T., Tarling, C. A., Withers, S. G., Henrissat, B., and Bourne, Y. (2004) Crystal structure of Thermotoga maritima alpha-L-fucosidase - Insights into the catalytic mechanism and the molecular basis for fucosidosis, *Journal of Biological Chemistry 279*, 13119-13128.

57. Watts, A. G., Oppezzo, P., Withers, S. G., Alzari, P. M., and Buschiazzo, A. (2006) Structural and kinetic analysis of two covalent sialosyl-enzyme intermediates on Trypanosoma rangeli sialidase, *Journal of Biological Chemistry 281*, 4149-4155.

58. Newstead, S. L., Potter, J. A., Wilson, J. C., Xu, G. G., Chien, C. H., Watts, A. G., Withers, S. G., and Taylor, G. L. (2008) The structure of Clostridium perfringens NanI sialidase and its catalytic intermediates, *Journal of Biological Chemistry 283*, 9080-9088.

59. Stoddart, J. F., In Stereochemistry of Carbohydrates; Wiley-Interscience:Toronto, Canada, 1971.

60. Davies, G. J., Ducros, V. M. A., Varrot, A., and Zechel, D. L. (2003) Mapping the conformational itinerary of beta-glycosidases by X-ray crystallography, *Biochemical Society Transactions 31*, 523-527.

61. Taylor, E. J., Goyal, A., Guerreiro, C. I. P. D., Prates, J. A. M., Money, V. A., Ferry, N., Morland, C., Planas, A., Macdonald, J. A., Stick, R. V., Gilbert, H. J., Fontes, C. M. G. A., and Davies, G. J. (2005) How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides - Structure and activity of a clostridium thermocellum lichenase, CtLic26A, *Journal of Biological Chemistry 280*, 32761-32767.

62. Jeffrey, G. A., and Taylor, R. (1980) The Application of Molecular Mechanics to the Structures of Carbohydrates, *Journal of Computational Chemistry 1*, 99-109.

63. Woods, R. J., Szarek, W. A., and Smith, V. H. (1991) A Comparison of Semiempirical and Abinitio Methods for the Study of Structural Features of Relevance in Carbohydrate-Chemistry, *Journal of the Chemical Society-Chemical Communications*, 334-337.

64. McNamara, J. P., Muslim, A. M., Abdel-Aal, H., Wang, H., Mohr, M., Hillier, I. H., and Bryce, R. A. (2004) Towards a quantum mechanical force field for carbohydrates: a reparametrized semi-empirical MO approach, *Chemical Physics Letters 394*, 429-436.

65. Ha, S. H., Gao, J. L., Tidor, B., Brady, J. W., and Karplus, M. (1991) Solvent Effect on the Anomeric Equilibrium in D-Glucose - a Free-Energy Simulation Analysis, *Journal of the American Chemical Society 113*, 1553-1557.

66. Barrows, S. E., Storer, J. W., Cramer, C. J., French, A. D., and Truhlar, D. G. (1998) Factors controlling, relative stability of anomers and hydroxymethyl conformers of glucopyranose, *Journal of Computational Chemistry 19*, 1111-1129.

67. Barrows, S. E., Dulles, F. J., Cramer, C. J., French, A. D., and Truhlar, D. G. (1995) Relative Stability of Alternative Chair Forms and Hydroxymethyl Conformations of Beta-D-Glucopyranose, *Carbohydrate Research 276*, 219-251.

68. Tvaroska, I., Andre, I., and Carver, J. P. (2000) Ab initio molecular orbital study of the catalytic mechanism of glycosyltransferases: Description of reaction pathways and determination of transition-state structures for inverting N-acetylglucosaminyltransferases, *Journal of the American Chemical Society 122*, 8762-8776.

69. Molteni, C., and Parrinello, M. (1997) Condensed matter effects on the structure of crystalline glucose, *Chemical Physics Letters 275*, 409-413.

70. Fernandez-Alonso, M. D., Canada, J., Jimenez-Barbero, J., and Cuevas, G. (2005) Theoretical study of inversion and topomerization processes of

substituted cyclohexanes: The relevance of the energy 3D hypersurface, *Chemphyschem 6*, 671-680.

71. Biarnes, X., Ardevol, A., Planas, A., Rovira, C., Laio, A., and Parrinello, M. (2007) The conformational free energy landscape of beta-D-glucopyranose. implications for substrate preactivation in beta-glucoside hydrolases, *Journal of the American Chemical Society 129*, 10686-10693.
72. Cremer, D., and Pople, J. A. (1975) General Definition of Ring Puckering Coordinates, *Journal of the American Chemical Society 97*, 1354-1358.
73. Guimaraes, B. G., Souchon, H., Lytle, B. L., Wu, J. H. D., and Alzari, P. M. (2002) The crystal structure and catalytic mechanism of cellobiohydrolase CelS, the major enzymatic component of the Clostridium thermocellum cellulosome, *Journal of Molecular Biology 320*, 587-596.
74. Guerin, D. M. A., Lascombe, M. B., Costabel, M., Souchon, H., Lamzin, V., Beguin, P., and Alzari, P. M. (2002) Atomic (0.94 angstrom) resolution structure of an inverting glycosidase in complex with substrate, *Journal of Molecular Biology 316*, 1061-1069.
75. Miyake, H., Kurisu, G., Kusunoki, M., Nishimura, S., Kitamura, S., and Nitta, Y. (2003) Crystal structure of a catalytic site mutant of beta-amylase from Bacillus cereus var. mycoides cocrystallized with maltopentaose, *Biochemistry 42*, 5574-5581.
76. Davies, G. S., M. L.; Withers, S. G. Glycosyl transfer. In comprehensive Biological Catalysis; Sinnott, M. L., Ed.; AcademicPress: 1998; Vol. 1, pp 119-208.
77. Wolfenden, R., Lu, X. D., and Young, G. (1998) Spontaneous hydrolysis of glycosides, *Journal of the American Chemical Society 120*, 6814-6815.
78. Asano, N. (2003) Glycosidase inhibitors: update and perspectives on practical use, *Glycobiology 13*, 93R-104.
79. Whitworth, G. E., Macauley, M. S., Stubbs, K. A., Dennis, R. J., Taylor, E. J., Davies, G. J., Greig, I. R., and Vocadlo, D. J. (2007) Analysis of PUGNAc and NAG-thiazoline as transition state analogues for human O-GlcNAcase: Mechanistic and structural insights into inhibitor selectivity

and transition state poise, *Journal of the American Chemical Society 129*, 635-644.

80. Tailford, L. E., Offen, W. A., Smith, N. L., Dumon, C., Morland, C., Gratien, J., Heck, M. P., Stick, R. V., Bleriot, Y., Vasella, A., Gilbert, H. J., and Davies, G. J. (2008) Structural and biochemical evidence for a boat-like transition state in beta-mannosidases, *Nature Chemical Biology 4*, 306-312.
81. Kawatkar, S. P., Kuntz, D. A., Woods, R. J., Rose, D. R., and Boons, G. J. (2006) Structural basis of the inhibition of Golgi alpha-mannosidase II by mannostatin A and the role of the thiomethyl moiety in ligand-protein interactions, *Journal of the American Chemical Society 128*, 8310-8319.
82. Mosi, R., Sham, H., Uitdehaag, J. C. M., Ruiterkamp, R., Dijkstra, B. W., and Withers, S. G. (1998) Reassessment of acarbose as a transition state analogue inhibitor of cyclodextrin glycosyltransferase, *Biochemistry 37*, 17192-17198.
83. Berland, C. R., Sigurskjold, B. W., Stoffer, B., Frandsen, T. P., and Svensson, B. (1995) Thermodynamics of Inhibitor Binding to Mutant Forms of Glucoamylase from Aspergillus-Niger Determined by Isothermal Titration Calorimetry, *Biochemistry 34*, 10153-10161.
84. Andersen, O. A., Nathubhai, A., Dixon, M. J., Eggleston, I. M., and van Aalten, D. M. F. (2008) Structure-based dissection of the natural product cyclopentapeptide chitinase inhibitor argifin, *Chemistry & Biology 15*, 295-301.
85. Tarling, C. A., Woods, K., Zhang, R., Brastianos, H. C., Brayer, G. D., Andersen, R. J., and Withers, S. G. (2008) The search for novel human pancreatic alpha-amylase inhibitors: High-throughput screening of terrestrial and marine natural product extracts, *Chembiochem 9*, 433-438.
86. Yuzwa, S. A., Macauley, M. S., Heinonen, J. E., Shan, X. Y., Dennis, R. J., He, Y. A., Whitworth, G. E., Stubbs, K. A., McEachern, E. J., Davies, G. J., and Vocadlo, D. J. (2008) A potent mechanism-inspired O-GlcNAcase inhibitor that blocks phosphorylation of tau in vivo, *Nature Chemical Biology 4*, 483-490.

87. Collins, P. J., Haire, L. F., Lin, Y. P., Liu, J. F., Russell, R. J., Walker, P. A., Skehel, J. J., Martin, S. R., Hay, A. J., and Gamblin, S. J. (2008) Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants, *Nature 453*, 1258-U1261.

88. Lee, J. K., Bain, A. D., and Berti, P. J. (2004) Probing the transition states of four glucoside hydrolyses with C-13 kinetic isotope effects measured at natural abundance by NMR spectroscopy, *Journal of the American Chemical Society 126*, 3769-3776.

89. Schramm, V. L. (2007) Enzymatic transition state theory and transition state analogue design, *Journal of Biological Chemistry 282*, 28297-28300.

90. Asano, N. (2003) Glycosidase inhibitors: update and perspectives on practical use, *Glycobiology 13*, 93r-104r.

91. Moscona, A. (2005) Drug therapy - Neuraminidase inhibitors for influenza, *New England Journal of Medicine 353*, 1363-1373.

92. Brameld, K. A., and Goddard, W. A. (1998) Substrate distortion to a boat conformation at subsite-1 is critical in the mechanism of family 18 chitinases, *Journal of the American Chemical Society 120*, 3571-3580.

93. Andre, G., Kanchanawong, P., Palma, R., Cho, H., Deng, X., Irwin, D., Himmel, M. E., Wilson, D. B., and Brady, J. W. (2003) Computational and experimental studies of the catalytic mechanism of Thermobifida fusca cellulase Cel6A (E2), *Protein Engineering 16*, 125-134.

94. Espinosa, J. F., Montero, E., Vian, A., Garcia, J. L., Dietrich, H., Schmidt, R. R., Martin-Lomas, M., Imberty, A., Canada, F. J., and Jimenez-Barbero, J. (1998) Escherichia coli beta-galactosidase recognizes a high-energy conformation of C-lactose, a nonhydrolizable substrate analogue. NMR and modeling studies of the molecular complex, *Journal of the American Chemical Society 120*, 1309-1318.

95. Biarnes, X., Nieto, J., Planas, A., and Rovira, C. (2006) Substrate distortion in the michaelis complex of Bacillus 1,3-1,4-beta-glucanase - Insight from first principles molecular dynamics simulations, *Journal of Biological Chemistry 281*, 1432-1441.

96. Moura-Tamames, S., Ramos, M. J., and Fernades, P. A. (2009) Modelling beta-1,3-Exoglucanase- saccharide interactions: structure of the enzyme-substrate complex and enzyme binding to the cell wall, *Febs Journal 276*, 156-157.

97. Thomas, A., Jourand, D., Bret, C., Amara, P., and Field, M. J. (1999) Is there a covalent intermediate in the viral neuraminidase reaction? A hybrid potential free-energy study, *Journal of the American Chemical Society 121*, 9693-9702.

98. Bottoni, A., Miscione, G. P., and De Vivo, M. (2005) A theoretical DFT investigation of the lysozyme mechanism: Computational evidence for a covalent intermediate pathway, *Proteins-Structure Function and Bioinformatics 59*, 118-130.

99. Bowman, A. L., Grant, I. M., and Mulholland, A. J. (2008) QM/MM simulations predict a covalent intermediate in the hen egg white lysozyme reaction with its natural substrate, *Chemical Communications*, 4425-4427.

100. Petersen, L., Ardevol, A., Rovira, C., and Reilly, P. J. (2009) Mechanism of Cellulose Hydrolysis by Inverting GH8 Endoglucanases: A QM/MM Metadynamics Study, *Journal of Physical Chemistry B 113*, 7331-7339.

101. Moura-Tamames, S. A., Ramos, M. J., and Fernades, P. A. (2009) Exploring beta-1,3-Exoglucanase-polysaccharide interactions: molecular dynamics of the enzyme-substrate complex with or without a polysaccharide ligand bound in the flexible enzyme loops, *Febs Journal 276*, 156-156.

102. Moura-Tamames, S. A., Ramos, M. J., and Fernandes, P. A. (2009) Modelling beta-1,3-exoglucanase-saccharide interactions: Structure of the enzyme-substrate complex and enzyme binding to the cell wall, *Journal of Molecular Graphics & Modelling 27*, 908-920.

103. Greig, I. R., and Williams, I. H. (2007) Glycosidase inhibitors as conformational transition state analogues, *Chemical Communications*, 4422-4422.

104. Whitfield, D. M. (2009) Computational studies of the role of glycopyranosyl oxacarbenium ions in glycobiology and glycochemistry, *Advances in Carbohydrate Chemistry and Biochemistry*, 62, 83-159.

3. Endo-β-1,4-xylanase from *Bacillus circulans* of family G/11(BCX)

Overview

In this chapter, I give a detailed description of the endo-β-1,4-xylanase from *Bacillus circulans* as a member of family G/11 xylanases, covering many general topics pressing upon those we have handled in this work. In this study I have exclusively investigated some aspects that have not been well-understood through experiments neither could be explained on a static level via X-ray crystallographic techniques. These aspects include distortion of the substrate sugar ring in the active site, ambiguities around the precise catalytic pathway, and mutagenic studies for some strategic residues indirectly involved in the catalytic landscape.

3.1. X-ray crystallographic data

A substantial volume of both structural and functional information is available on the family G/11 xylanases (*1*). The three-dimensional fold of many members of this family has been determined, and the two catalytic residues implicated in the double-displacement mechanism have been identified through a combination of sequence analysis, mutational studies, inhibition experiments, and structure determination (*2-7*).

A member of the family G (*8, 9*) or family 11 xylanases (*10*), *Bacillus circulans* xylanase (BCX), is a relatively small retaining glycosidase (20 kDa), for which an X-ray crystal structure has been reported (*4, 11*).

3.1.1. The outcome of the X-ray analysis

3.1.1.1. Catalytic residues and mechanism of action

The summary of the results that came out of the previously mentioned studies (*4, 11*) indicated that the enzyme operates via general acid catalysis using a nucleophile (or charge-stabilizing residue) and an acid-base catalyst with retention of the substrate's anomeric configuration adopting double displacement mechanism (see chapter 2). The active site of this enzyme includes two glutamic

acid residues Glu78 and Glu172 that are intimately involved in catalysis; this feature is absolutely conserved in this family of xylanases. Arg112 also plays a role in catalysis, and Tyr69 and Tyr80 are likely to be involved in substrate binding (Fig.3.1). (However, see chapters 5 and 6 for more insight regarding the role of Tyr69.)

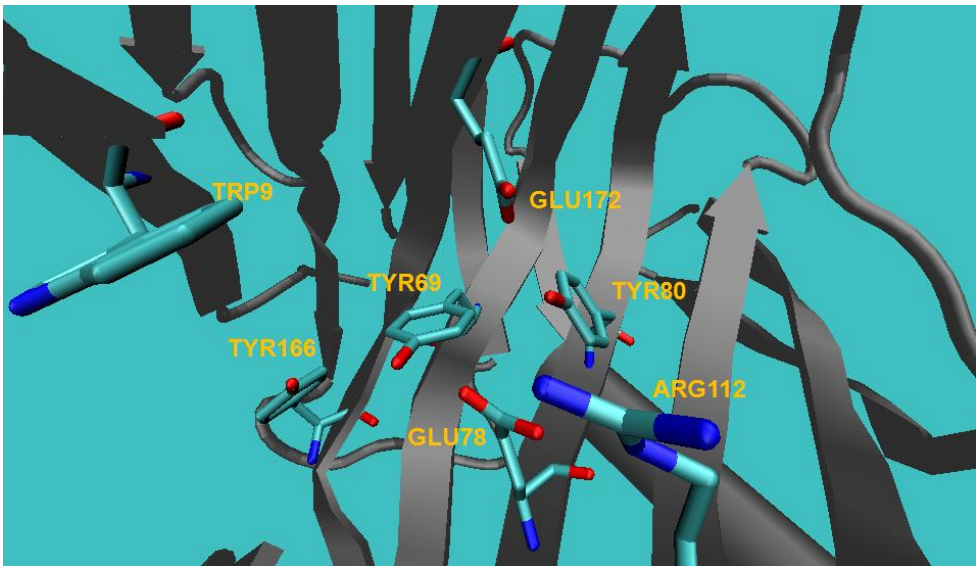


Fig.3.1. The active site of the B. circulans xylanases. The structure of the enzyme has been presented using the molecular graphic program VMD

Mutational analysis combined with X-ray crystallographic studies of an enzyme-substrate complex have provided a more detailed description of the active site and the mechanism of action for this enzyme (*4*). In the following text, I sum up the main points that have evolved from these studies:

The specific activities of the acidic residue mutants showed that only mutation of Glu78, and Glu172 have a significant effect on the catalysis (Table 3.1). When glutamic acid residues are mutated to glutamine (E78Q and E172Q), no enzyme activity can be detected; however, when the carboxylate side chain is maintained but the side chain length is shortened in E78D and E172D, some residual activity is observed, 0.04% and 0.16% for E78D and E172, respectively (Table 3.1). This dramatic decrease in the activity, strongly suggests that that Glu78 and Glu172 are the catalytic amino acid residues.

In addition to the conserved acidic residues, Glu78 and Glu172, there are a number of other conserved residues in the catalytic cleft (Fig.3.1). Most

importantly, Tyr69 and Tyr80, Tyr166 and Arg112 are thought to take part in the catalytic scheme. Mutation of the conserved Tyr80 to phenylalanine (Y80F) resulted in a dramatic loss of the activity, leaving only 0.03% residual activity (Table 3.1), whereas, the mutation Y69F resulted in a complete loss of detectable enzyme activity. In contrast to the previous findings, the mutation Y166F has a minor effect on the enzyme functionality. Hence the contribution of the Tyr166 hydroxyl groups may not be that important. Mutation of the conserved Arg112 to lysine (R112K) also resulted in no change in the enzyme activity, but mutation to asparagine (R112N) resulted in a 68% drop in specific activity.

Table.3.1. Enzymatic properties of various xylanases mutants *(4)*.

Enzyme	Activity (% of wild type)	K_m (mg/mL)	K_{cat} (min^{-1})	k_{cat}/k_m (min^{-1} mg^{-1})
BCX (wild type)	100	1.9 ± 0.2	8,918	4,693
E78D	0.043	2.3 ± 0.2	3.9	1.7
E78Q	< 0.01	nd	nd	nd
E172D	0.19	4.4 ± 0.5	17.1	3.9
E172Q	< 0.01	nd	nd	nd
E172C	< 0.01	nd	nd	nd
Y69 F	< 0.01	nd	nd	nd
Y80F	0.045	12 ± 2	3.7	0.3
Y166F	76	1.2 ± 0.12	6,807	5,672
R112K	83	5.9 ± 0.4	7,373	1,249
R112N	35	5.5 ± 0.7	3,122	568

Supporting evidence for the role of Glu78 as the nucleophile is the proximity of Arg112 to Glu78. Despite the absence of salt link in this structure, the presence of a positive charge in the vicinity may help to decrease the pK_a of Glu78, thus keeping it as a negatively charged species. Moreover, the catalytic activity is more affected by E78D than E172D, which is consistent with E78 being more crucial as a nucleophile.

As concluded from the mutation analysis studies, tyrosines in the active site of the *Bacillus circulans* xylanases are implicated in the catalytic mechanism. However, their role is still uncertain. In the structure of this enzyme complexed with xylotetraose, hydroxyl groups of the xylose rings are close enough to form

hydrogen bonds with the hydroxyl group of Tyr69 and Tyr166. In this study, the authors (*4*) reason the complete loss of the enzyme activity in the case of Y69F, but not Y166F, as that Tyr69 makes two hydrogen bonds, one with the C2OH of the second (non-reducing) xylose ring, and the second with OE1 atom of Glu78. The hydrogen bond with Glu78 may help orient the nucleophile for the reaction, whereas the second one helps to position the substrate. For the enzyme activity then, the single hydrogen bond between Tyr166 and C3OH of the second xylose ring seems to be less important.

However these studies provided a detailed description of enzyme architecture, still some key features could not be explained precisely on that static level of the X-ray crystal structures. For instance, the role of Tyr69 on the catalytic course is still unconvincing.

In the same directions, and the following text, I provide more recent studies that could provide more understandings of the *Bacillus circulans* xylanases. Most of these studies are based upon some kinetic analysis.

3.2. Trials towards more understanding

Since the experiments done by Campbell et al on *Bacillus circulans* xylanases (*4*), a vast number of studies including kinetic measurements has been carried out to uncover the points of ambiguity for this enzyme and provide better understanding on the catalytic pathway as well as to probe the function of the critical active site residues implicated in the catalytic landscape.

3.2.1. Conformation of the covalently bonded disaccharide and the neighbouring active site residues

Kinetic and spectroscopic studies have been done on a glycosyl-enzyme intermediate trapped using the mechanism-based inhibitor 2',4'-dinitrophenyl 2-deoxy-2-fluoro-β-xylobioside (DNPFXb) (Fig.3.2) (*7, 12*). The electronegative fluorine atom at 2-position of the xylose moiety slows both the formation and hydrolysis of the intermediate by inductively destabilizing the oxacarbenium ion-like transition states and by eliminating an important hydrogen bonding interaction at this position (*13, 14*). The xylosyl-enzyme intermediate is formed

with a second-order rate constant, k_i/K_i of 0.34 $min^{-1}mM^{-1}$ at 40 ºC and pH 6.0 (*7*). The intermediate then slowly hydrolyzes with $t_{1/2}$ = 350 min under the same conditions.

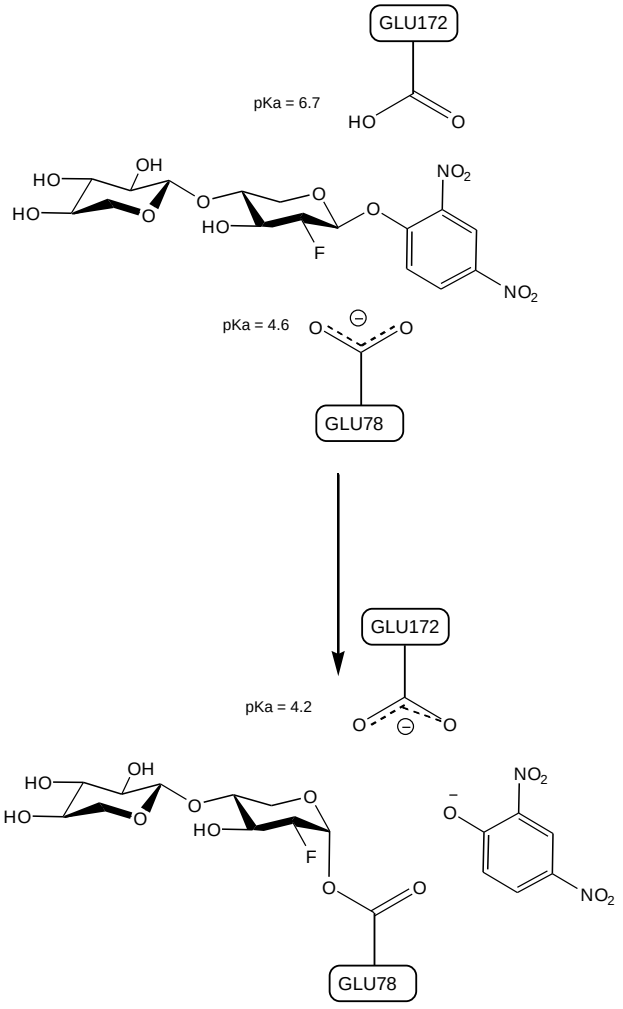


Fig.3.2. Schematic representation of the formation of the xylosyl-enzyme intermediate species using DNPFXb as mechanism-based inhibitor.

3.2.1.1. Conformation of the covalently bonded disaccharide

The structure of the BCX-2FXb Glycosyl-enzyme intermediate was determined to a resolution of 1.8 Å (*15*) (Fig.3.3). In comparison to the free enzyme as determined by Campbell et al (*5*), the final model for BCX-2FXb Glycosyl-enzyme intermediate was being best defined and exhibits excellent stereochemistry.

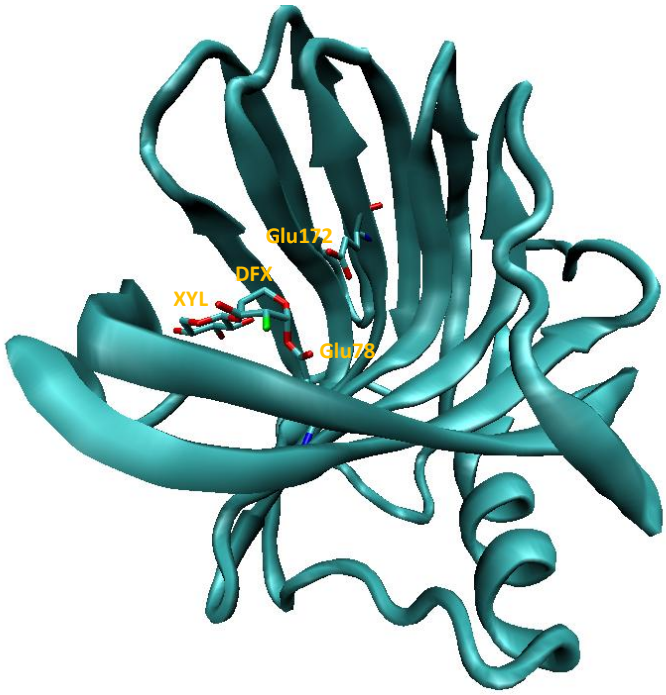


Fig.3.3. The three-dimensional structure of BCX-2FXb glycosyl-enzyme intermediate along with the two glutamate residues implicated in the double-displacement mechanism. The freely distributed graphical program VMD has been used to create this 3D structure.

The X-ray structure of enzyme-intermediate shows the 2-deoxy-2-fluoro-xylobiose (2FXb) ligand covalently bonded to the nucleophilic residue, Glu78, via an α-anomeric linkage with the C1 atom of the proximal sugar ring *syn* to the newly formed ester group. The proximal sugar ring occupies the -1 subsite of BCX. The bond distance between the C1 of the proximal sugar ring and $O^{\epsilon 2}$ atom of Glu78 is 1.45 Å. This observation confirms that the formation of glycosyl-enzyme intermediate happens through a double-displacement mechanism (Fig.3.2).

The xylose moiety covalently bonded to Glu78 is heavily distorted from the conventional $^{4}C_{1}$ (chair) conformation and adopts a $^{2,5}B$ (boat) conformation instead (Fig.3.4). This latter conformation allows C5, O5, C1 and C2 atom of the proximal sugar ring to achieve a nearly planar geometry (0.05 Å rms deviation from planarity versus 0.22 Å rms deviation for the same atoms in the distal saccharide which adopts $^{4}C_{1}$ conformation).

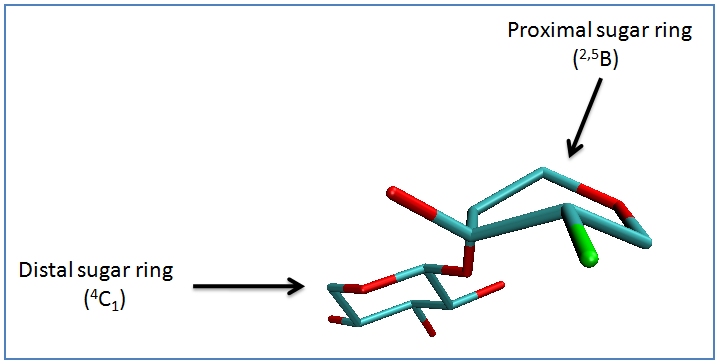


Fig.3.4. A diagram showing the proximal sugar ring adopting ^{2,5}B conformation, while the distal saccharide adopts ^{4}C$_1$ conformation. Covalent bond with the enzyme is removed for clarity.

As mentioned earlier, glycosylation and deglycosylation steps of the reaction catalyzed by BCX operate via transition states with substantial oxacarbenium ion-like character in which a partial double bond connecting O5 and C1. This requires C5, O5, C1 and C2 to approach coplanarity at the oxacarbenium ion-like TS. In this study, the authors proposed that the ^{2,5}B conformation could evolve from ^{4}C$_1$ conformation via ^{4}C$_1$→ ^{2}H$_3$ → ^{2}S$_0$ → ^{2,5}B itinerary (*15*). If that is the case, so the movement is confined mostly to atoms C5 and O5, and the dihedral angle defined by atoms C1, C2, C3 and C4 is not required to undergo much change (this angle is observed to be 48º in the proximal saccharide and 57º in the distal saccharide). As the atoms C5 and O5 of xylose are not impeded by the functional groups found on C1, C2, C3 and C4 and thus not much energetic cost is expected when these atoms, C5 and O5, experience the movement described by the abovementioned itinerary.

3.2.1.2. The neighbouring active site residues

The X-ray crystallography for BCX-2FXb glycosyl-enzyme intermediate (*15*) reveals that the acid/base catalyst, Glu172, occupies a similar position and conformation in the wild-type BCX structure (*5*). O$^{\epsilon 2}$ atom of Glu172, which donates and abstracts protons during the catalytic course, forms a 3.10 Å hydrogen bond to a water molecule which lies 3.92 Å away from the C1 atom of the proximal sugar ring (Fig.3.5). This water molecule is also hydrogen bonded to the phenolic oxygen atom of Tyr80 (d = 2.74 Å), as well as to a second water molecule (d = 3.17 Å) and is a likely to be the one implicated in the deglycosylation step of the reaction. pK_a measurements showed that the O$^{\epsilon 2}$ atom

of Glu172 is deprotonated under the experimental conditions (pK_a = 4.2) and hence it must accept a hydrogen bond from a water molecule in the vicinity (*12*, *16*). Tyr80, which appears to be donating a hydrogen bond (*d* = 3.83 Å) to the deprotonated $O^{\epsilon 2}$ atom of Glu172, must then also accept a hydrogen bond from the water molecule. This hydrogen bond network allows one of the lone pairs of the tetrahedral oxygen atom of this water molecule to be directed towards the second solvent molecule while the other lone pair points in the general direction of the proximal saccharide. Such configuration is ideal for an attack on the C1 of proximal saccharide during the deglycosylation step (Fig.3.5).

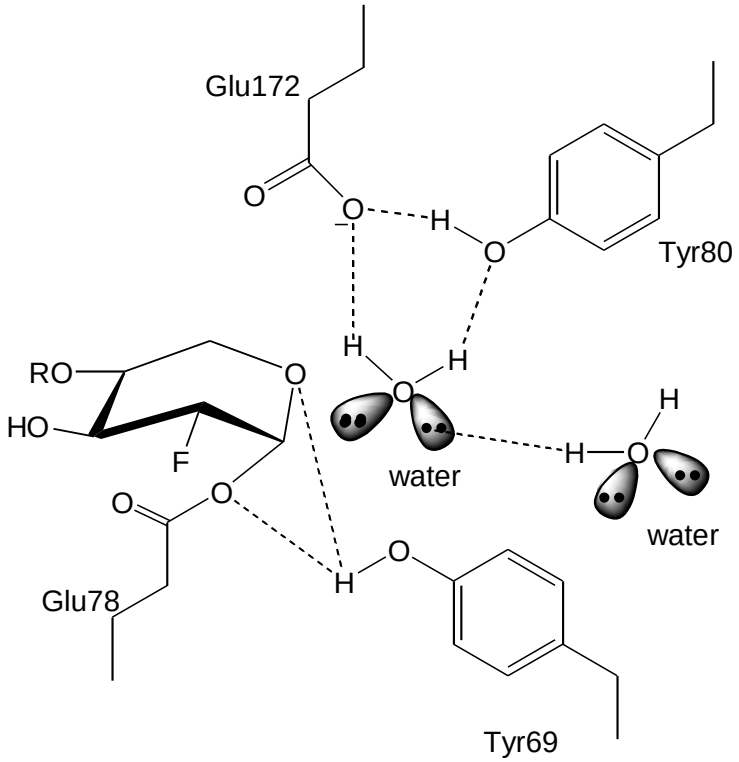


Fig.3.5. A schematic representation of the hydrogen bond network in the active site for BCX-2FXb Glycosyl-enzyme intermediate showing mainly the solvent molecule expected to act as a nucleophile in the deglycosylation step.

The role of Tyr69 in the catalysis has been paid much interest for BCX-2FXb Glycosyl-enzyme intermediate as well. What has been newly seen in the BCX-2FXb intermediate, compared to what had been found by Campell *et al* (*4*), is that the phenolic oxygen of Tyr69 in BCX-2FXb forms a stronger interaction with the endocyclic oxygen (O5) of the proximal sugar moiety than with the $O^{\epsilon 2}$ atom of Glu78 (*d* = 2.99 Å). The nature of this interaction is fascinating since

the hydroxyl group of Tyr69 is very important for the catalysis, as evident by the fact that the Y69F variant of BCX exhibits no detectable enzyme activity (*4*). It has been hypothesized that Tyr69 plays a crucial role in positioning Glu78, and the loss of the hydrogen bond between Tyr69 and Glu78 in Y69F BCX would alter the position of the nucleophilic carboxylate, Glu78, and thereby disrupt catalysis (*15*). Since the crystal structure of the Y69F variant of BCX has been solved at 1.5 Å resolution (*15*), this hypothesis was no longer rational as the position of Glu78 remains essentially unperturbed in the case of Y69F structure. Therefore, the precise placement of Glu78 must be dictated by other interactions, such as those with the nearby Gln127, and the role of Tyr69 may be hidden somehow else.

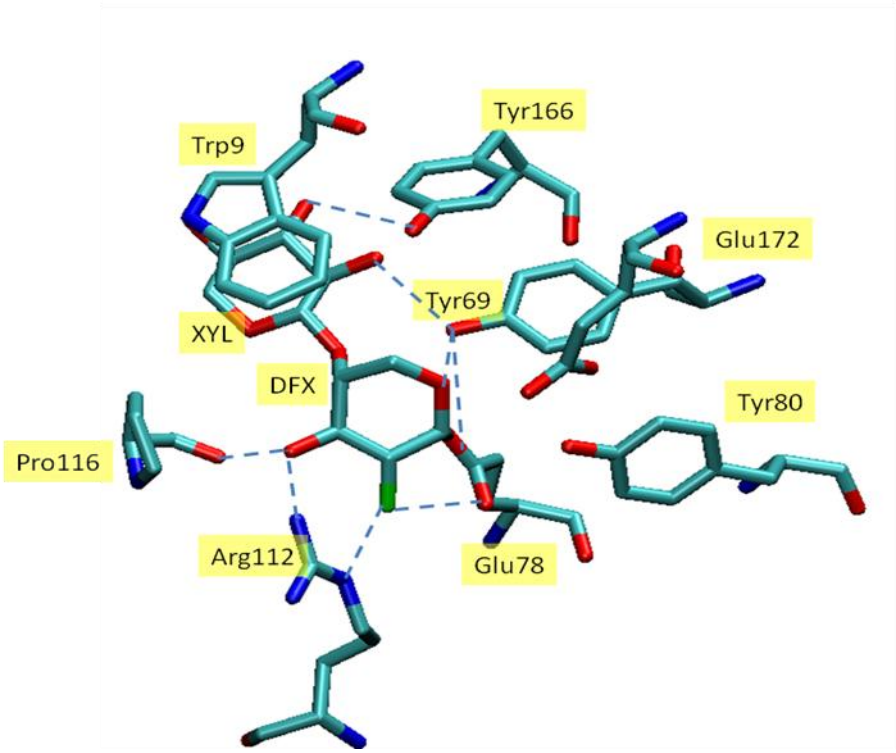


Fig.3.6. A perspective view showing interactions of the 2FXb disaccharide with the nearby active site residues

A detailed analysis of the hydrogen bond geometry about Tyr69 suggests that the hydrogen bond formed between Tyr69 and the endocyclic oxygen (O5) atom of the proximal saccharide could stabilize the ^{2,5}B conformation (*15*). Tyr69 could stabilize the oxacarbenium ion-like TS via a dipolar interaction between the phenolic oxygen and the endocyclic oxygen (O5), and that happens at the TS as

Tyr69 preferentially favours the hydrogen bond formation with the negatively charged $O^{\epsilon 2}$ atom of Glu78.

3.2.3. Positioning the catalytic residues

In retaining glycosidases, including xylanases, the two catalytic carboxylic acid residues, Glu78 and Glu172, are positioned ~5.5 Å apart in the active site. In the inverting glycosidases, which employ a direct displacement mechanism, these residues are typically further apart (~10-11 Å) (*17*). This distance between the two essential carboxylic acids therefore seems to be essential for controlling the mechanism, however it is unlikely to be the only criterion.

One of the interesting studies that have been achieved to probe the effect of altering the distance between the catalytic residues, Glu78 and Glu172, on the enzyme catalysis has unveiled some important facts around this subject (*18, 19*). The strategy adopted for altering the distance between Glu78 and Glu172 involved selectively replacing each glutamic acid with its shortened, aspartic acid, or lengthened analogue, carboxymethylated cysteine. Detailed kinetic evaluation has shown that the replacement of Glu172 by a group with no significant capacity as a proton donor/acceptor (Gln or Cys) results in an enzyme which cannot cleave xylan, a natural substrate which requires acid catalysis (*4*). However, these mutants were able to hydrolyze aryl xylobiosides with relatively good leaving groups which need little acid catalysis. The k_{cat}/K_m values for this experiment reflects the first irreversible step, glycosylation, were decreased only 8-30 fold for substrate of aglycon pK_a , 5.5. Greater difference were observed for substrates of higher pK_a, with no hydrolysis being detected for phenyl xylobioside (pK_a = 9.99) (*18*). The presence of carboxylic group at 172 position, even if improperly positioned, xylan hydrolysis rates are much closer to those obtained for the wild-type enzyme, the k_{cat} values for carboxymethylated-Glu172Cys and Glu172Asp being reduced only 25 and 400-fold, respectively. Interestingly, this difference between carboxymethylated-Glu172Cys and Glu172Asp essentially disappears when the enzyme is assayed with the aryl xylobiosides, presumably because of their decreased need for protonic assistance. Rates measured for these two enzymes were generally greater than for the Gln

and Cys mutants which reflects the fact that a mispositioned acid catalyst is better that none at all (*18*).

Replacement of Glu78, the active site nucleophile, with a cysteine had a dramatic effect on the enzymatic activity, the apparent k_{cat} value for the Glu78Cys mutant being decreased at least 7×10^6-fold relative to native xylanases. Whether the drop in the catalytic activity in the case of Glu78Cys is due to the inability of Cys78 to act as active site nucleophile/leaving group or because Cys78 is not close enough to reach the substrate's anomeric carbon still needs more precise investigation. However, carboxymethylated-Glu78Cys exhibited much improved enzymatic activity with a range of aryl xylobioside substrates. Values of k_{cat} are reduced only some 3-10 fold relative to the native xylanases. A covalent glycosyl-enzyme intermediate has been detected for the carboxymethylated-Glu78Cys, thus a retaining mechanism still proceeds. On the other hand, studies on Glu78Asp mutant implies that shortening the nucleophile side chain reduce the catalytic activity by 1600 – 500 fold relative to wild-type enzyme (*18*).

To summarize the aforementioned results, In *B.circulans* xylanases, a proton transfer group at position 172 is essential for the hydrolysis of substrates that need acid catalysis, such as phenyl xylobiosides and the natural substrate, xylan. However such a group is not essential for substrates with good leaving groups ($pK_a < 5.5$). Removal of such a group decreased k_{cat} value by only 50 – 125 fold relative to the native enzyme. Shortening or lengthening of this carboxyl side chain at site 172 results, in approximately the same little decrease in k_{cat}/K_m values (1000 fold for xylan, and 2 – 23 fold for aryl xylobiosides).

In contrast to the position 172, shortening of the nucleophilic side chain (Glu78Asp) decreased k_{cat}/K_m values at least 1600 fold for the aryl xylobiosides and much more for xylan, whereas extending the side chain (carboxymethylated-Glu78Cys) decreased k_{cat}/K_m values by 16 – 100 fold. Thus, the positional requirements for proton transfer are less demanding than those for nucleophilic attack.

3.3. Beyond experiments

Although a huge amount of work has been done to provide more insight on the structural and mechanistic picture of *B.circulans* xylanases, there are many

points that could not be explained precisely on the static level of the experimental basis. The fact is that enzymes are actually dynamic molecules which interact in numerous ways to simply catalyze the conversion of reactants into products. Thus enzymes must be able to change conformation in order to function properly. During this dynamic itinerary the interaction patterns change from one state to another along the reaction course. In light of this fact, the capability of molecular modelling to simulate the dynamic state of the enzymes as well as the possible mechanistic routes could provide clear understanding for many aspects which cannot be probed experimentally.

In this thesis, and via a diverse number of computational techniques, we provided a dynamic picture for many aspects. In particular, ring distortion and the energetic cost associated with this feature to occur, a detailed description of the interactions within the active site pocket over the time scale of the reaction pathway, deeper understanding of possible mechanistic route, and some mutational modelling that helped understand some key features of this enzyme.

3.4. Applications on xylanases

Biologically, xylanases play a major role in the digestive system of herbivorous micro-organisms (mammals, conversely, do not produce xylanase). Additionally, xylanases are present in fungi for the degradation of plant matter into usable nutrients *(20)*.

Commercially, there has been much industrial interest in xylan and its hydrolytic enzymatic complex, as a supplement in animal feed, for the manufacture of bread, food and drinks, textiles, bleaching of cellulose pulp, ethanol and xylitol production *c*

In the future, xylanase may be used for the production of biofuel from unusable plant material *(20)*.

References

1. Torronen, A., and Rouvinen, J. (1997) Structural and functional properties of low molecular weight endo-1,4-beta-xylanases, *J. Biotechnology*, **57**, 137-149.
2. Torronen, A., and Rouvinen, J. (1995) Structural comparison of two major endo-1,4-xylanases from *Trichoderma reesei*, *Biochemistry*, **34**, 847-856.
3. Krengel, U., and Dijkstra, B. W. (1996) Three-dimensional structure of endo-1,4-beta-xylanase I from *Aspergillus niger*: molecular basis for its low pH optimum, *J. Mol. Biol.*, **63**, 70-78.
4. Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J., and Yaguchi, M. (1994) Mutational and crystallographic analyses of the active-site residues of the *Bacillus circulans* xylanase, *Protein Sci.*, **3**, 467-475.
5. Wakarchuk, W., Methot, N., Lanthier, P., Sung, W., Seligy, V., Yaguchi, M., To, R., Campbell, R., and Rose, D. (1992) The 20-kD xylanase of *Bacillus subtilis* - a structure-function Analysis, *Xylans and Xylanases*, **7**, 439-442
6. Withers, S. G., and Aebersold, R. (1995) Approaches to labeling and identification of active-site residues in glycosidases, *Protein Sci.*, **4**, 361-372.
7. Miao, S. C., Ziser, L., Aebersold, R., and Withers, S. G. (1994) Identification of glutamic-acid-78 as the active-site nucleophile in *Bacillus subtilis* xylanase using electrospray tandem mass-spectrometry, *Biochemistry*, **33**, 7027-7032.
8. Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., and Warren, R. A. J. (1991) Domains in microbial beta-1,4-glycanases - sequence conservation, function, and enzyme families, *Microbiological Rev.*, **55**, 303-315.
9. Oku, T., Roy, C., Watson, D. C., Wakarchuk, W., Campbell, R., Yaguchi, M., Jurasek, L., and Paice, M. G. (1993) Amino-acid-sequence and thermostability of xylanase-a from *Schizophyllum commune*, *FEBS Lett.*, **334**, 296-300.

10. Henrissat, B., and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino-acid-sequence similarities, *Biochem. J.*, **293**, 781-788.

11. Gebler, J., Gilkes, N. R., Claeyssens, M., Wilson, D. B., Beguin, P., Wakarchuk, W. W., Kilburn, D. G., Miller, R. C., Warren, R. A. J., and Withers, S. G. (1992) Stereoselective hydrolysis catalyzed by related beta-1,4-glucanases and beta-1,4-xylanases, *J. Biol. Chem.*, 267, 12559-12561.

12. McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996) The pK_a of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a C-13-NMR study of *Bacillus circulans* xylanase, *Biochemistry,* **35**, 9958-9966.

13. Withers, S. G., Street, I. P., Bird, P., and Dolphin, D. H. (1987) 2-Deoxy-2-fluoroglucosides - a novel class of mechanism-based glucosidase inhibitors, *J. Am. Chem. Soc.*, **109**, 7530-7531.

14. Withers, S. G., Rupitz, K., and Street, I. P. (1988) 2-Deoxy-2-fluoro-D-glycosyl fluorides - a new class of specific mechanism-based glycosidase inhibitors, *J. Biol. Chem.*, **263**, 7929-7932.

15. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase, *Biochemistry,* **38**, 5346-5354.

16. Joshi, M. D., Hedberg, A., and McIntosh, L. P. (1997) Complete measurement of the pK_a values of the carboxyl and imidazole groups in *Bacillus circulans* xylanase, *Protein Sci.*, **6**, 2667-2670.

17. Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases, *Structure,* **3**, 853-859.

18. Lawson, S. L., Wakarchuk, W. W., and Withers, S. G. (1997) Positioning the acid/base catalyst in a glycosidase: studies with *Bacillus circulans* xylanase, *Biochemistry,* **36**, 2257-2265.

19. Lawson, S. L., Wakarchuk, W. W., and Withers, S. G. (1996) Effects of both shortening and lengthening the active site nucleophile of *Bacillus circulans* xylanase on catalytic activity, *Biochemistry,* **35**, 10110-10118.

20. Collins, T., Gerday, C., and Feller, G. (2005) Xylanases, xylanase families and extremophilic xylanases, *FEMS Microbiology Rev.*, **29**, 3-23.

4. Work Objectives, choice of a crystal structure.

4.1. Objectives

Despite the availability of a substantial number of experiments that provide us with a lot of structural and functional information on the family G/11 xylanases (see chapter 2), many questions still remain concerning details of the reaction pathway and catalytic mechanism need to be precisely answered. Most importantly, why does the proximal sugar-ring of the substrate distort from $^{4}C_{1}$ chair to $^{2,5}B$ boat conformation? Why is Tyr69 so crucial for the enzyme catalysis, as evidenced by the fact that mutation at this site leads to a complete loss of the enzyme activity? What is the preferred mechanistic pathway? And finally, how important is the proper positioning of the two catalytic residues, Glu78 and Glu172, in the active site? In light of these questions, I split up the rest of this thesis in such a way each question is answered in a separate chapter.

The following sketch sums up the points of interest that this project is based upon for *Bacillus circulans* xylanases (BCX) (Fig.4.1).

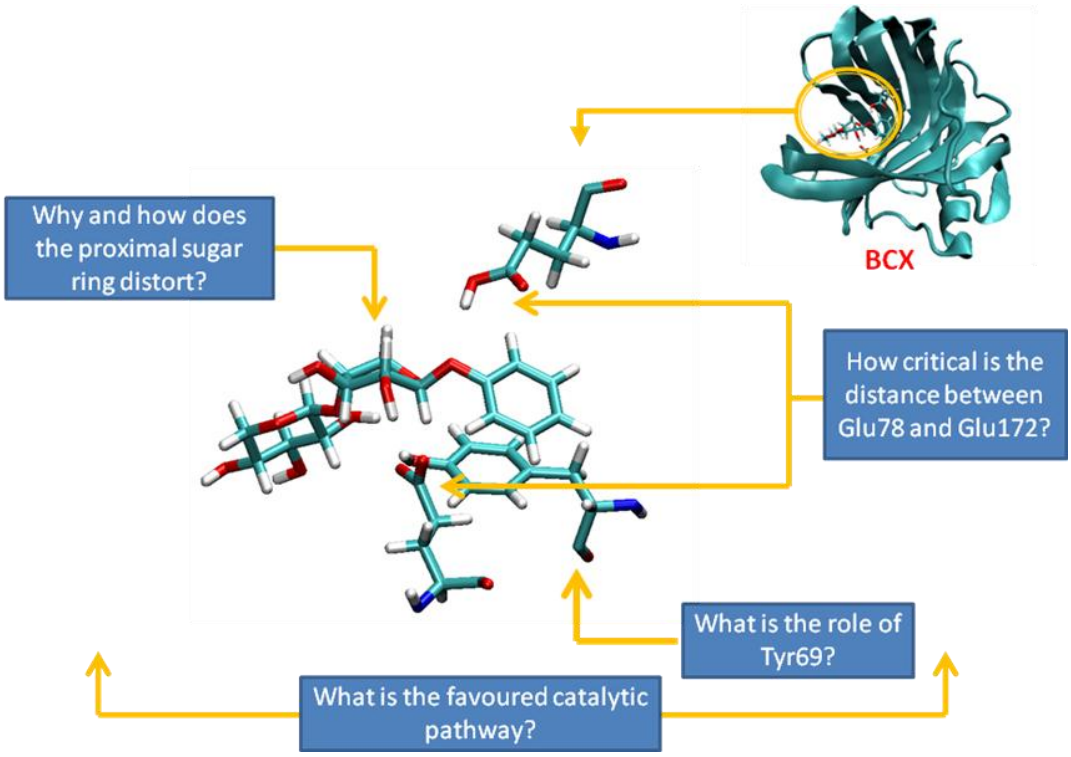


Fig.4.1. A sketch illustrating the objectives of the project.

For the sake of clarity, I make each chapter as a separate unit. In other words, each of the following chapters handles a particular point of interest, relevant methodology and technical procedures, results, discussion and conclusions.

4.2. Choice of an appropriate crystal structure

The choice of a well-defined crystal structure to start off with was not an easy task. In other words, the crystal structure should fulfil certain criteria that allow us to investigate the aforementioned points shown in figure 4.1. To this end and after a careful examination of all crystal structures determined for family G/11 xylanases from *Bacillus circulans*, I establish a table that summarize some key features for all resolved PDB structures for this family.

4.2.1. A survey of resolved BCX structures

There are about 10 resolved crystal structures for BCX/GH11 (see http://www.cazy.org/fam/GH11.html). We have looked precisely into these structures to choose a reasonable and well-resolved configuration to start with along this project (Fig.4.2 and Table.3.2).

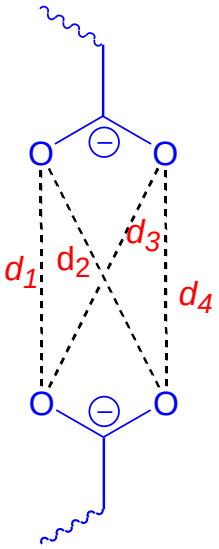


Fig.4.2. A sketch showing distances measured between the catalytic residues.

Table.4.1. Some key features along with the measured distances between the two catalytic residues for 10 members of retaining BCX.

PDB code	Distance between the catalytic residues				Key features
	d1	*d2*	*d3*	*d4*	
1BCX*(1)*	7.29	8.06			- Resolution 1.8 Å - Mutated -Non-covalent substrate exists (xylobiose) - Monomer
1BVV*(2)*	5.57	5.97	5.79	6.18	- Resolution 1.8 Å - wild enzyme - DFX-XYL-covalently bonded disaccharide exists - Monomer
1C5H(*3*)	5.59	6.40	6.43	5.61	- No substrate in the active site - Mutated - Monomer
1C5I(*3*)	6.08	5.59	6.24	5.50	- Mutated - Substrate (XYS-DFX) covalently bounded - Monomer
1HV0(*4*)	5.49	6.44	5.64	6.32	- No ligand in the active site - Monomer - Mutated
1HV1(*4*)	5.63	5.86	6.48	6.63	- No ligand - Monomer - Mutated
1XNB(*5*)	5.29	6.22	5.57	6.33	- No ligand - Monomer
1XNC(*5*)	5.26	6.26	5.48	6.35	- No ligand - Monomer
2BVV(*6*)	5.75	6.19	6.15	6.41	- No ligand - Monomer

In light of the data in this table, we considered 1BVV to be an optimum structure to start off with as no mutations have been done in any part of this enzyme

structure. Also, the substrate is covalently-bonded with the enzyme and exhibits a good stereochemical configuration. On top of that, it is well-defined structure and resolved at reasonable resolution. Also, we have measured the distance between the two catalytic residues over the whole series to take these values as a criterion, along with many others, which should be monitored along our computational simulations.

References

1. Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J., and Yaguchi, M. (1994) Mutational and crystallographic analyses of the active-site residues of the *Bacillus circulans* xylanase, *Protein Sci.*, **3**, 467-475.
2. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase, *Biochemistry*, **38**, 5346-5354.
3. Joshi, M. D., Sidhu, G., Pot, I., Brayer, G. D., Withers, S. G., and McIntosh, L. P. (2000) Hydrogen bonding and catalysis: A novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase, *Journal of Molecular Biology 299*, 255-279.
4. oshi, M. D., Sidhu, G., Nielsen, J. E., Brayer, G. D., Withers, S. G., and McIntosh, L. P. (2001) Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase, *Biochemistry 40*, 10115-10139.
5. Wakarchuk, W., Methot, N., Lanthier, P., Sung, W., Seligy, V., Yaguchi, M., To, R., Campbell, R., and Rose, D. (1992) The 20-Kd Xylanase of Bacillus-Subtilis - a Structure-Function Analysis, *Xylans and Xylanases 7*, 439-442
6. Wakarchuk, W. W., Sung, W. L., Campbell, R. L., Cunningham, A., Watson, D. C., and Yaguchi, M. (1994) Thermostabilization of the Bacillus-Circulans Xylanase by the Introduction of Disulfide Bonds, *Protein Engineering 7*, 1379-1386.

5. Substrate Ring Distortion in BCX

5.1. A Brief Introduction

As we have mentioned in chapter 2 and 3, ring distortion is a common feature among glycosidases. In BCX the substrate sugar ring at subsite -1 adopts $^{2,5}B$ (boat), rather than $^{4}C_{1}$ (chair) conformation, to achieve the coplanarity of C5, O5, C1 and C2 which is essential at the oxacarbenium ion like TS. But what has not yet been stated in the literature is how could the enzyme stabilize this unfavourable distorted conformation in the active site pocket and what is the energetic cost to achieve that?

In this chapter we answer this question via a sequential scenario of computational modelling approaches.

5.2. Computational Methods and Technical Details

Enzyme-substrate initial structure

The X-ray crystallographic coordinates (Fig. 5.1a) of the covalent enzyme-inhibitor complex of wild-type BCX (PDB accession code 1BVV) (*1*) were modified as follows:

(1) pK_a values for amino-acid residues at the pH optimum 5.7 were estimated using the H^{++} programme (*2*).

(2) Hydrogen atoms were added in accordance with the predicted protonation state.

(3) The whole system (16475 atoms) was energy minimized to a residual gradient of less than 0.001 kJ mol^{-1} $Å^{-1}$ with a QM/MM method using the DYNAMO package (*3*). The QM region (68 atoms) comprised the covalently-attached substrate - a disaccharide of xylose (XYL) and 2-deoxyfluoroxylose (DFX) - and the two catalytic residues, Glu78 and Glu172, and was described by AM1 semi-empirical Hamiltonian (*4*) The MM region contained the rest of the enzyme, and was described by the OPLS-AA potential (*5, 6*). AM1/OPLS is commonly used in QM/MM studies of enzymic systems; although other semiempirical QM methods might give better results for small molecules *in*

vacuo, their use in conjunction with incompatible QM/MM non-bonded interaction parameters could detrimentally affect their performance for condensed-phase applications. QM link atoms were placed along the $C_\alpha C_\beta$ bonds of Glu78 and Glu172 (Fig. 5.1c). At this stage no additional water molecules were present beyond the crystallographic waters.

(4) A gas-phase QM/MM MD simulation (NVT, 300 K, 5 ps) was performed to pre-equilibrate the protein and to allow the substrate to accommodate in the binding cavity.

(5) The F atom at position 2 of the proximal sugar moiety of the covalently-attached substrate was changed to OH to simulate the natural substrate, xylose; the system was QM/MM energy-minimized and equilibrated by MD once again.

(6) The covalent bond between the anomeric carbon of the (α-anomer of the) attached substrate and the oxygen of Glu78 was broken, and a phenoxy (OPh) leaving group was inserted manually to form the β-anomer. This change required care to retain the integrity of the enzyme configuration prior to the whole system being freely optimized again and equilibrated by MD, still without solvating waters.

(7) The whole system was enveloped in a cubic box of TIP3P water (*7*) of side-length 55.5 Å. First, all water molecules were relaxed with a gradient minimizer, while keeping the protein structure frozen; then the system was subjected to a short MD pre-equilibration using mild constraining forces to maintain the desired interactions between the substrate and the catalytic residues. Next, the whole system was equilibrated by MD for 20 ps at 150 K, using the NVT ensemble, still with the frozen protein. Finally, the whole system was freely minimized without any constraints and subsequently equilibrated for 20 ps at 300 K.

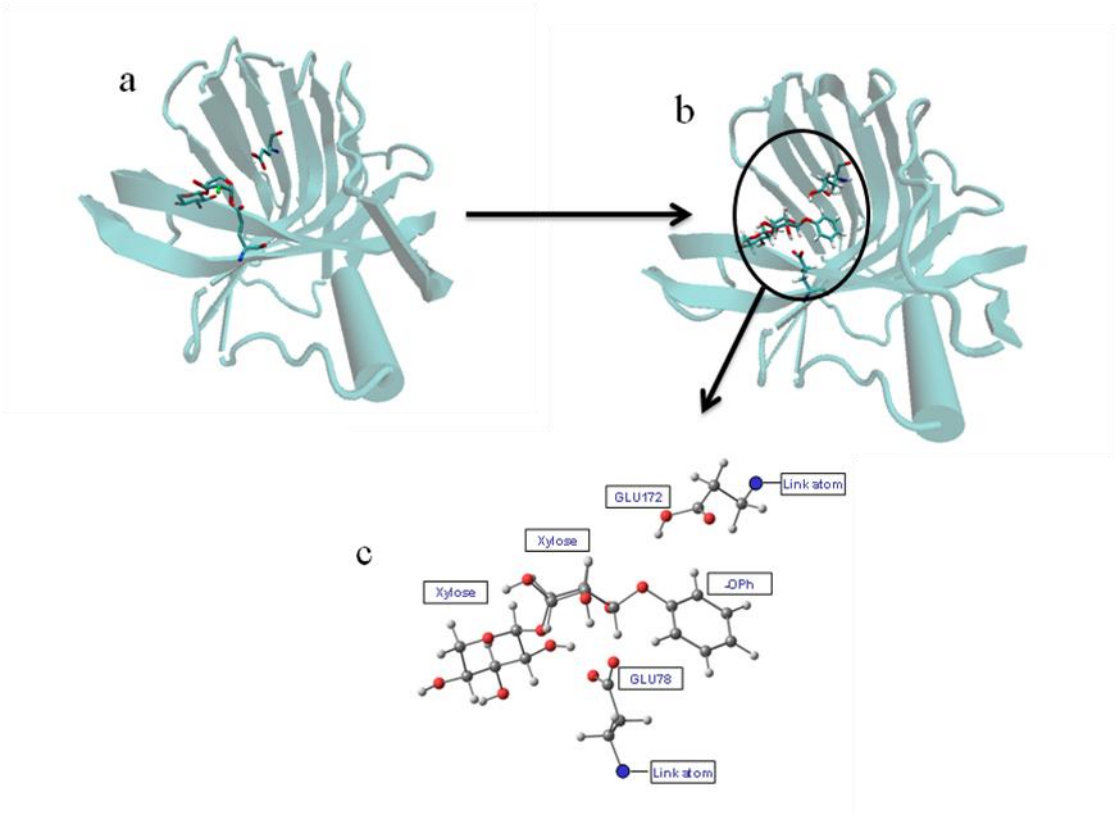


Fig. 5.1 (a) X-ray structure for wild-type BCX. (b) Final modelled structure after energy minimization and equilibration; water molecules are not shown for clarity. (c) QM atoms selected in the simulation.

Ring Distortion

In order to understand the nature of ring distortion at the −1 subsite within the environment of the enzyme active site, it is necessary also to explore the conformational behavior of the substrate in solution. To this end, we consider not only the final full system, as described in step 7 above, as model 3, but also two simpler models (Fig. 5.2): model 1 includes in the QM region only the proximal sugar ring of substrate in MM water, whereas model 2 additionally includes propionic acid and propionate groups in the QM region, to represent the functionality of Glu78 and Glu172; models 1 and 2 both involve cubic boxes of TIP3P water with side-length 31.4 Å. Subscripts B and C below denote the $^{2,5}B$ boat and $^{4}C_1$ chair conformations, respectively. The QM region of models 1 and 2 does not involve any link atoms.

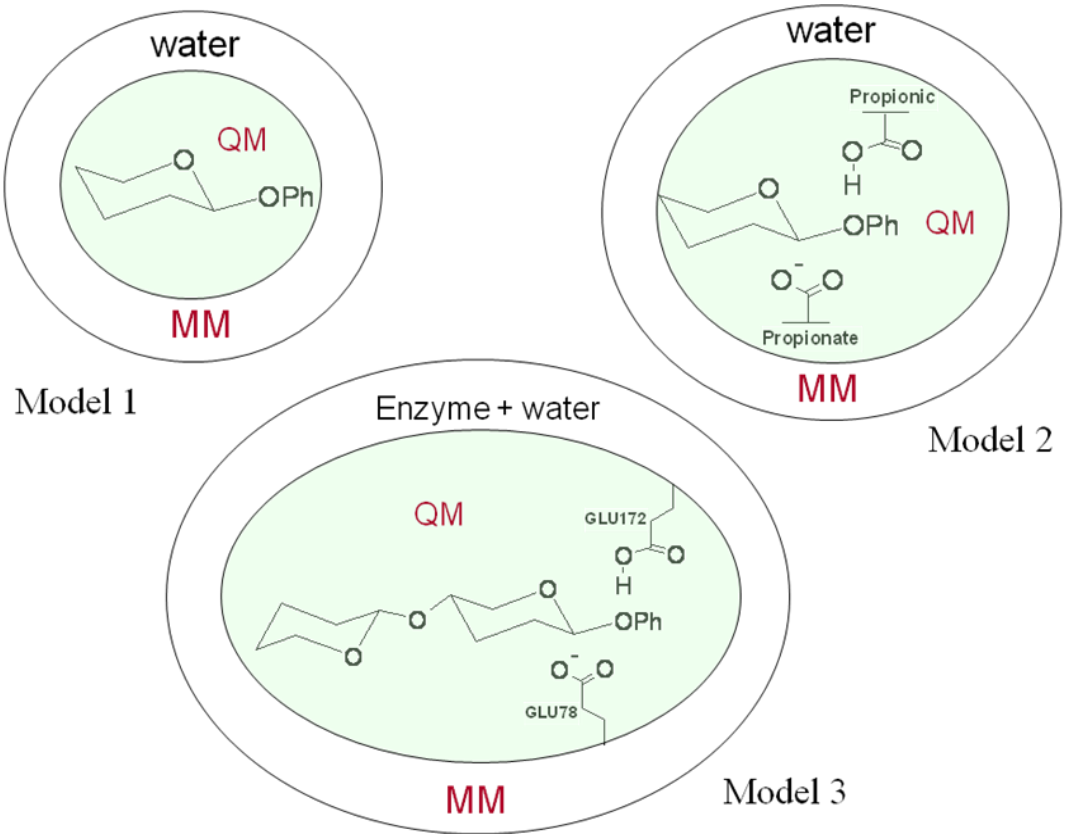


Figure 5.2 A schematic presentation of Model 1, model 2 and model 3, the ring conformation is not illustrated for feasibility.

Models 1 and 2 were allowed to equilibrate for 10 ps, and model 3 for 20 ps (same as above) of MD at 300 K in the NVT ensemble, prior to 30 ps of production MD for each model (1 fs timestep, switched cutoff radius of 16 Å applied to all interactions). No constraints were applied at any stage of the dynamic runs. Atomic charges were monitored along the dynamic simulation. A free-energy pathway for interconversion of $^{2,5}B$ and $^{4}C_1$ conformers was generated using the potential of mean force (PMF) approach and the weighted histogram analysis method (WHAM) as implemented in DYNAMO. Before running PMF calculations, it was necessary to determine the best internal coordinate to describe this conformational change. Ideally the conformations of hexapyranose rings may be described by combinations of distances and angles, such as the puckering coordinates of Cremer and Pople (*8*) (Fig.5.3) used in a recent study of glucose (*9*) but for simplicity we use a single coordinate. Five different dihedral angles were considered and the C3-C4-C5-O5 dihedral was

found to be the best for the purpose. An umbrella constraint force of 0.5 kJ mol^{-1} $Å^{-2}$ was applied in each of the 31 windows; 1 ps of equilibration was performed prior to 10 ps of production MD in each window, starting with a structure perturbed in the distinguished coordinate from the final structure in the previous window. Block averaging within each window suggests that the PMF is satisfactorily converged for its purpose in this study, although this does not mean that equilibrium has been achieved with respect to all hydroxyl group conformations.

Hexapyranose ring conformations were determined from trajectories for each model and were described following the approach of Bérces *et al (10)* Briefly, a conformation was specified in terms of a polar angle θ ($0° \leq \theta \leq 180°$), an equatorial angle ϕ ($0° \leq \phi < 360°$), and an amplitude *r*. Variation in the polar angle specifies a $^1C_4$ chair ($\theta = 0°$), envelope/half-chair (*E*/*H*, $\theta = 45°$), boat/skew-boat (*B*/*S*, $\theta = 90°$), envelope/half-chair (*E*/*H*, $\theta = 135°$), and $^4C_1$ chair ($\theta = 180°$) transformation; whereas variation in the equatorial angle specifies a boat/skew-boat (for $\phi = 90°$) or envelope/half-chair (for $\phi = 45°$or $135°$) pseudorotational itinerary. Angles θ and ϕ are sufficient to specify which of the 38 canonical conformations a molecule is closest to, while the conformational amplitude *r* specifies the extent of distortion from the planar cyclohexane structure. A diagrammatic representation of hexopyranose ring conformations in terms of θ and φ only is shown in Fig. 5.4.

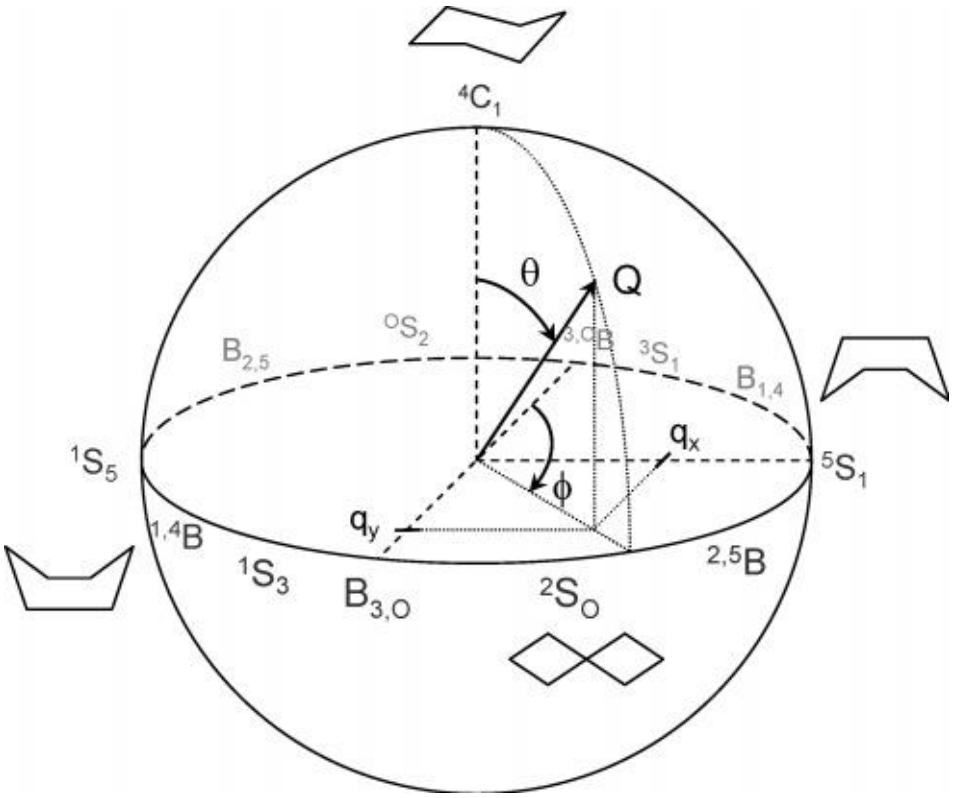


Fig.5.3 Cremer and Pople puckering coordinates of a six-membered ring (*10*).

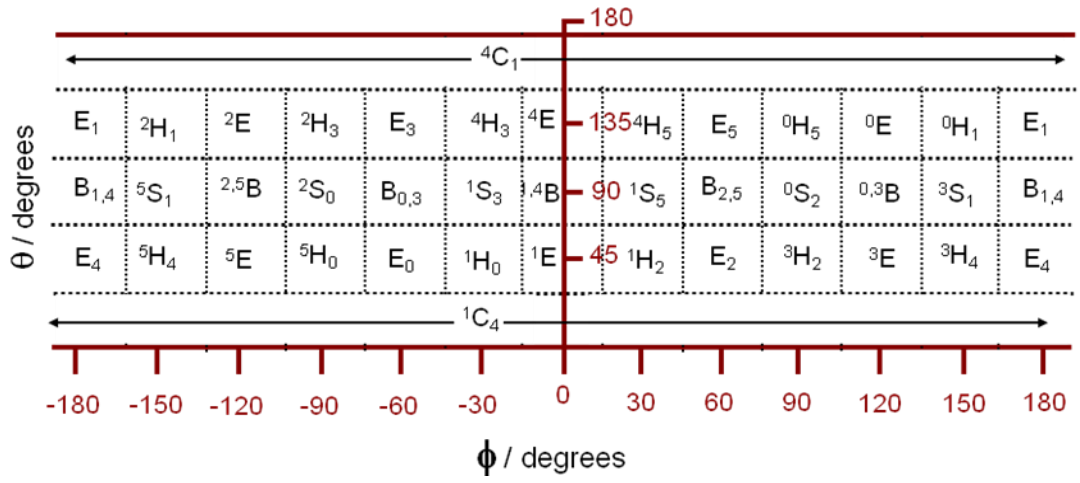


Fig. 5.4 Division of hexopyranose conformational space into regions corresponding to canonical conformations (*10*).

Optimization of Isolated substrates

Unconstrained geometry optimization was performed for the substrate, starting from each of the two conformations ($^{2,5}B$ and $^{4}C_1$), using the B3LYP density functional method[23] and 6-31+G(d) basis set, as implemented in the Gaussian03 package (*11*).

5.3 Results

Isolated substrates: model 1

B3LYP/6-31+G(d) geometry optimization starting from the $^{4}C_{1}$ conformer of the phenyl xyloside substrate in vacuum confirms the chair as a stable minimum-energy species. However, a similar optimization starting from the $^{2,5}B$ conformer leads to the $^{1}S_{3}$ skew conformation (Fig. 5.5) as a local minimum, demonstrating that the boat is not a stable conformer in the gas phase.

The QM/MM MD trajectory (Fig. 5.6a) initiated from the $^{2,5}B$ conformation of phenyl β-xyloside in water follows an itinerary through $B_{3,0}$ to $^{2}S_{0}$ and $^{2,5}B$ over the course of 30ps, as shown by the plot of ϕ vs. time; these conformations are clustered in the region bounded by $70° < \theta < 110°$ and $-120° < \phi < -45°$ (Fig. 5.6b). The corresponding trajectory initiated from the $^{4}C_{1}$ conformation in water remains stable: the apparent fluctuations in Fig. 5.6c and the spread of points in Fig. 5.6d reflect the ill-defined character of the angle ϕ when $\theta \approx 180°$.

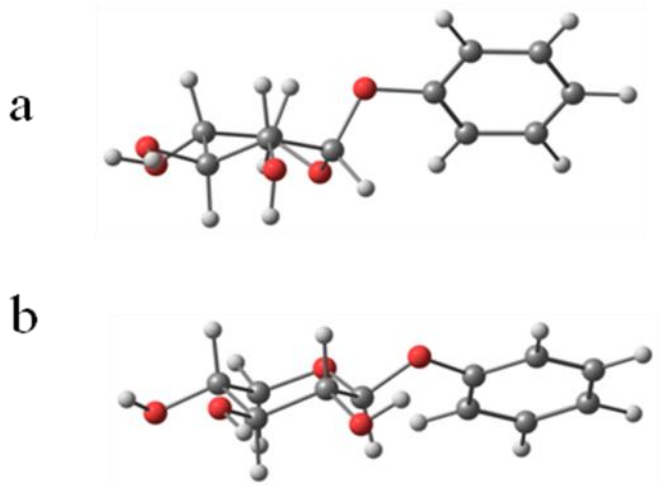
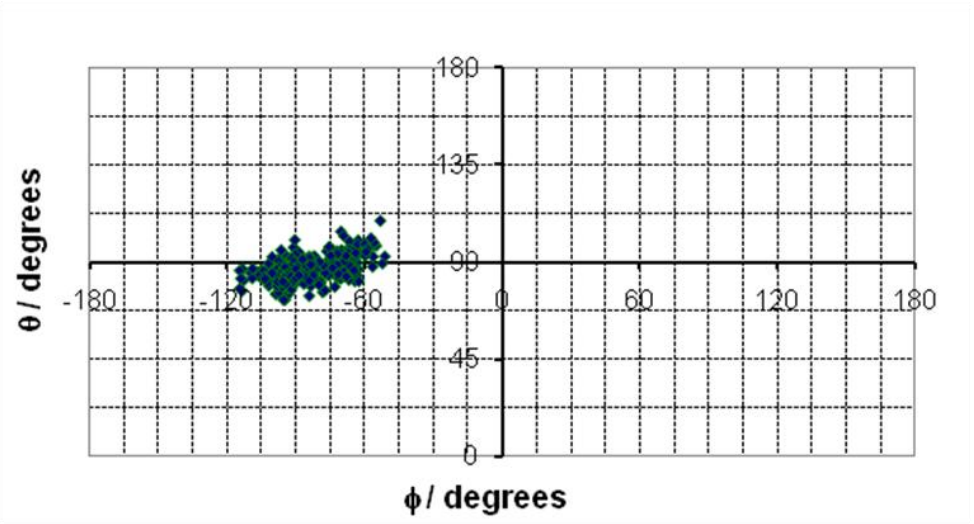
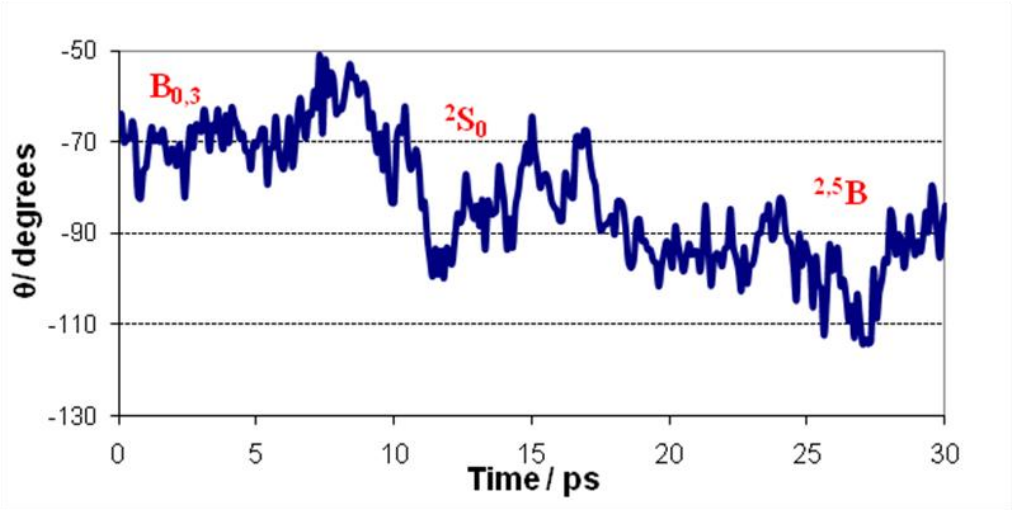


Fig.5.5 B3LYP/6-31+G(d) optimized structures starting from (a) $^{2,5}B$ and (b) $^{4}C_{1}$ conformations, respectively.

a


180
135
90
45
0
-180
-120
-60
0
60
120
180
θ / degrees
ϕ / degrees

b


-50
-70
-90
-110
-130
0
5
10
15
20
25
30
θ/ degrees
Time / ps
B0,3
2S0
2,5B

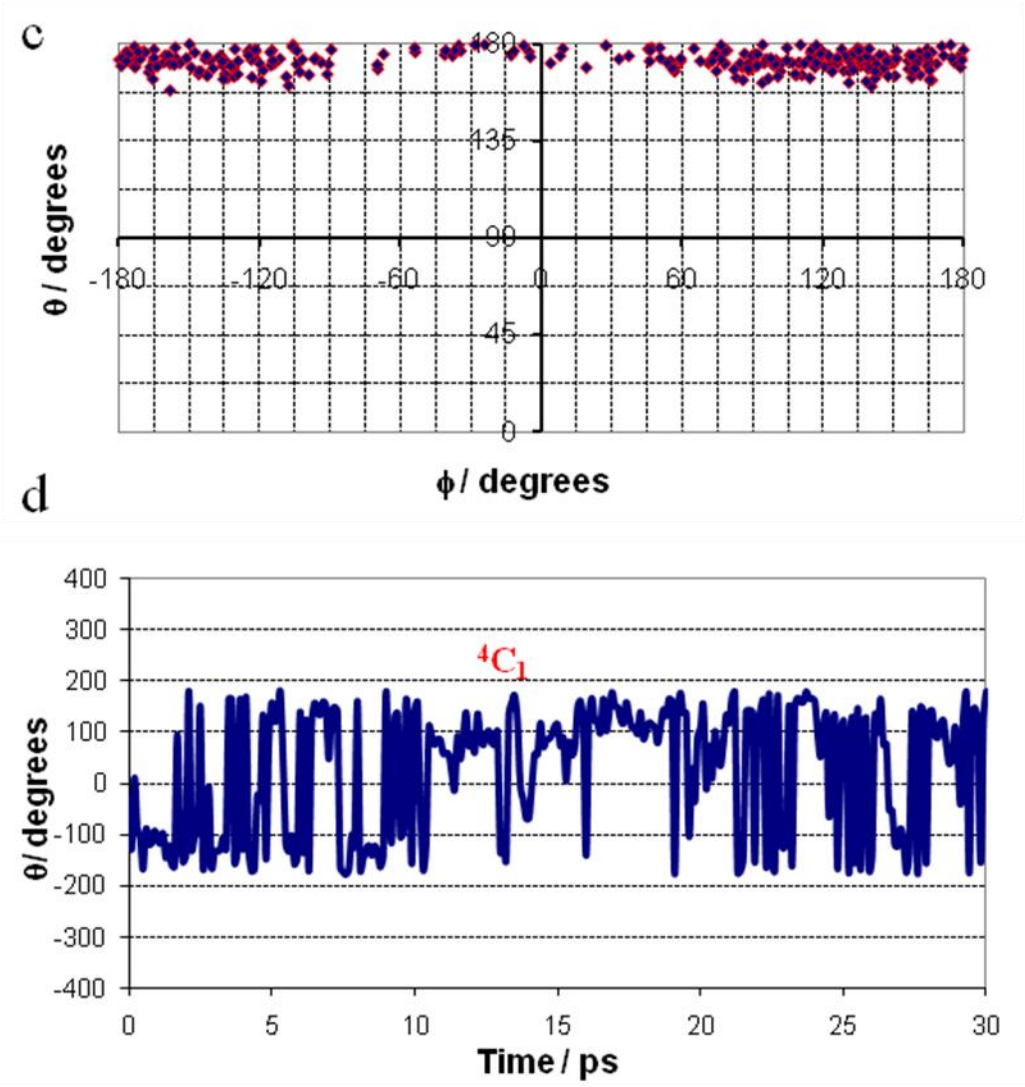


Fig. 5.6 QM/MM MD simulation for Model 1 in water: (a) and (c) plot of composite angle ϕ *vs*. time during the QM MD trajectory for xylose ring conformation starting from $^{2,5}B$ and $^4C_1$ conformations respectively; (b) and (d) conformational space sampled during the QM MD trajectory for $^{2,5}B$ and $^4C_1$ conformations respectively.

Effect of catalytic residues: model 2

The effects of the acid Glu172 and nucleophilic Glu78 residues may be mimicked simply by propionic acid and propionate moieties, respectively. The $^{2,5}B$ conformation with the additional groups located in positions taken from the structure of the enzyme-substrate complex optimizes to a local $^{2,5}B$ boat minimum in both gas-phase B3LYP/6-31+G(d) and QM/MM aqueous-phase calculations.

The QM/MM MD trajectory initiated from the $^{4}C_{1}$ conformation in water remains stable and behaves very similarly to that for model 1. The corresponding trajectory initiated from the $^{2,5}B$ boat in water fluctuates between this conformation and the $^{2}S_{0}$ skew-boat during 30ps, as shown by the plot of ϕ vs. time (Fig. 5.8a); these conformations are clustered in the region bounded by $75° < \theta < 105°$ and $-135° < \phi < -90°$ (Fig. 5.8b). In comparison with model 1, the observed relative stability of $^{2,5}B$ boat may refer to the stabilization by hydrogen bond between the OH group at C2 and the nucleophilic propionate residue (Fig. 5.7).

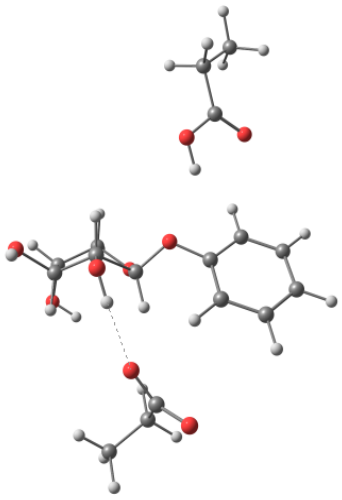
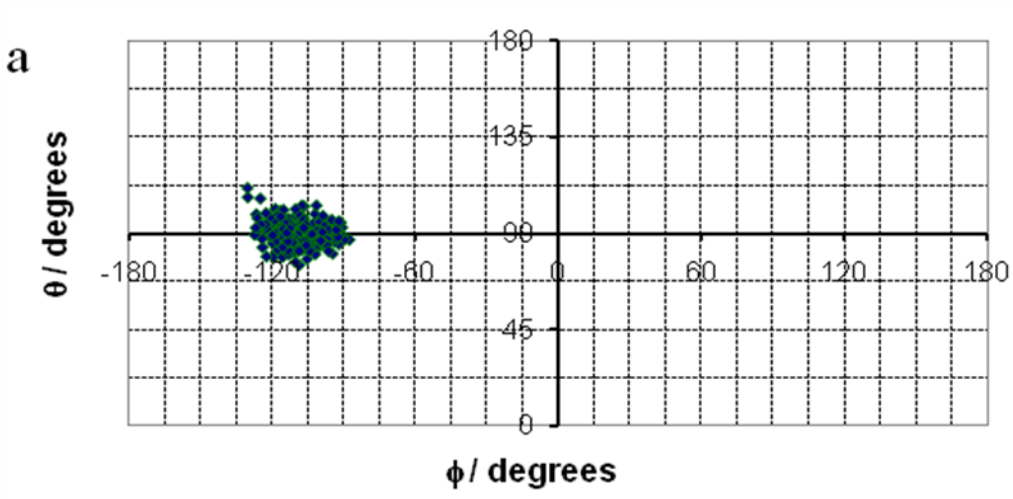


Fig. 5.7 Hydrogen bond stabilization between the ring OH at C2 and propionate.



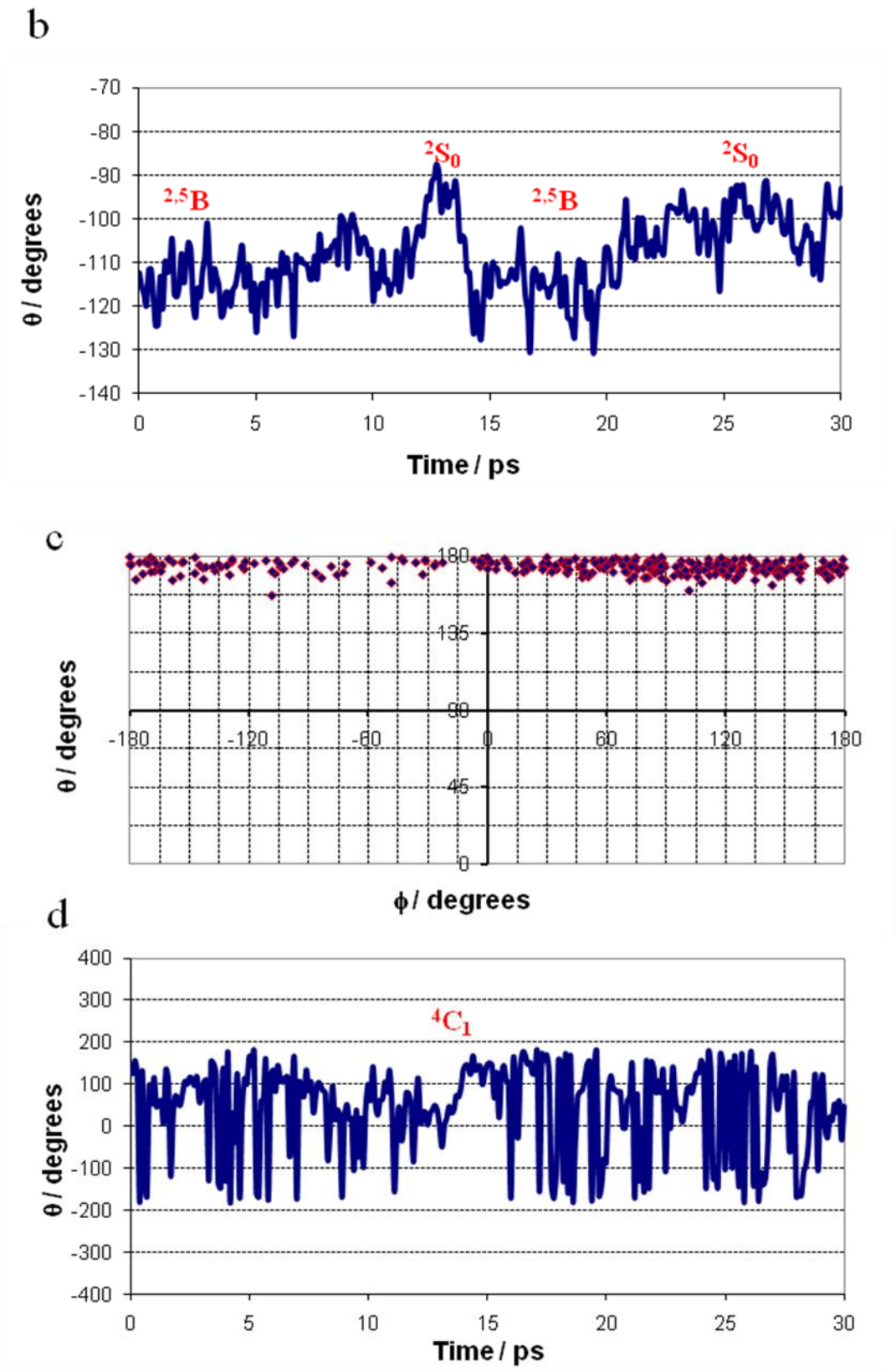
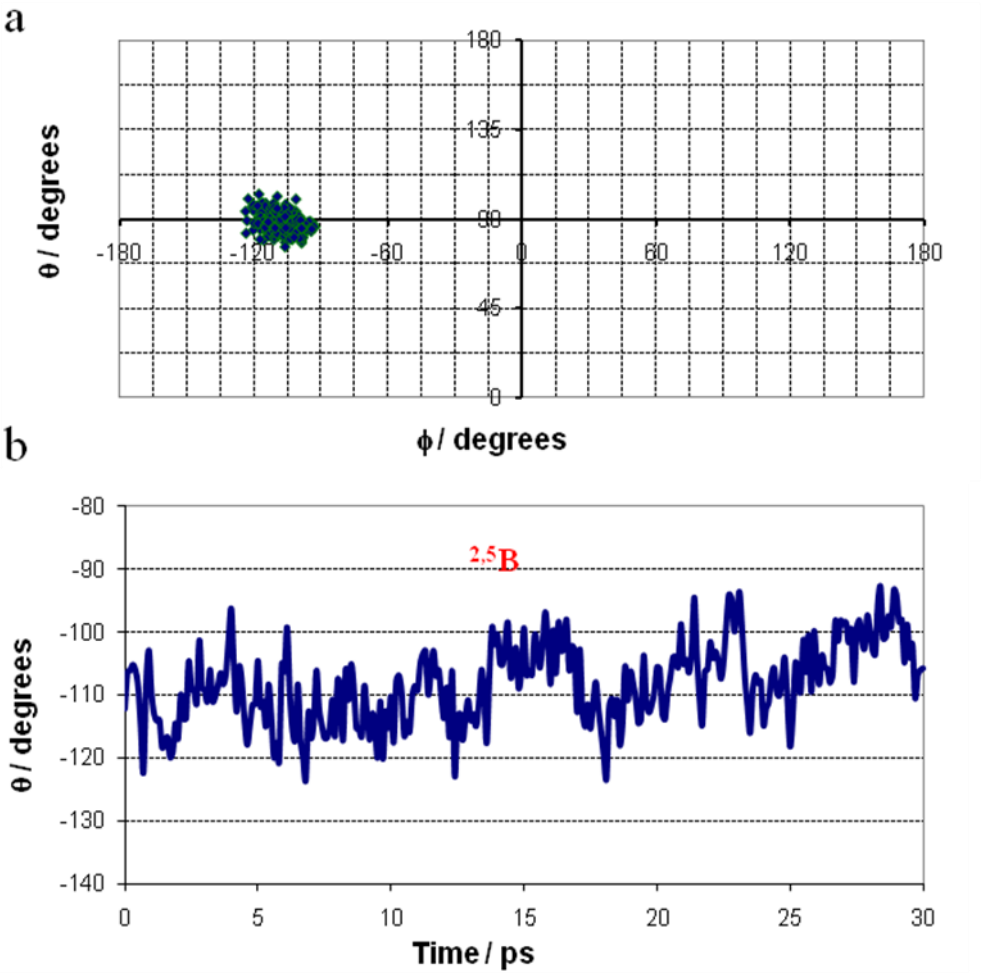


Fig. 5.8 QM/MM MD simulation for Model 2 in water: (a) and (c) plot of composite angle ϕ *vs*. time during the QM MD trajectory for xylose ring conformation starting from $^{2,5}B$ and $^{4}C_1$ conformations respectively; (b) and (d) conformational space sampled during the QM MD trajectory for $^{2,5}B$ and $^{4}C_1$ conformations respectively.

Effect of protein environment: model 3

The QM/MM MD trajectory (Fig. 5.9a and b) initiated from the $^{2,5}B$ conformer shows this to be a stable species which does not evolve towards a skew conformation. The C2-C1-O5-C5 dihedral angle of the $^{2,5}B$-substrate in the enzyme achieves a better coplanarity (0-30^{o}) than in model 2.

Interestingly, after initial energy minimization, the $^{4}C_1$ chair conformation of the proximal sugar ring within the enzyme active was not found as a local minimum but converted to an E_1 envelope and then transformed quickly to $^{2,5}B$ (via $^{2}H_1$ and ^{2}E) within 5 ps of MD simulation.



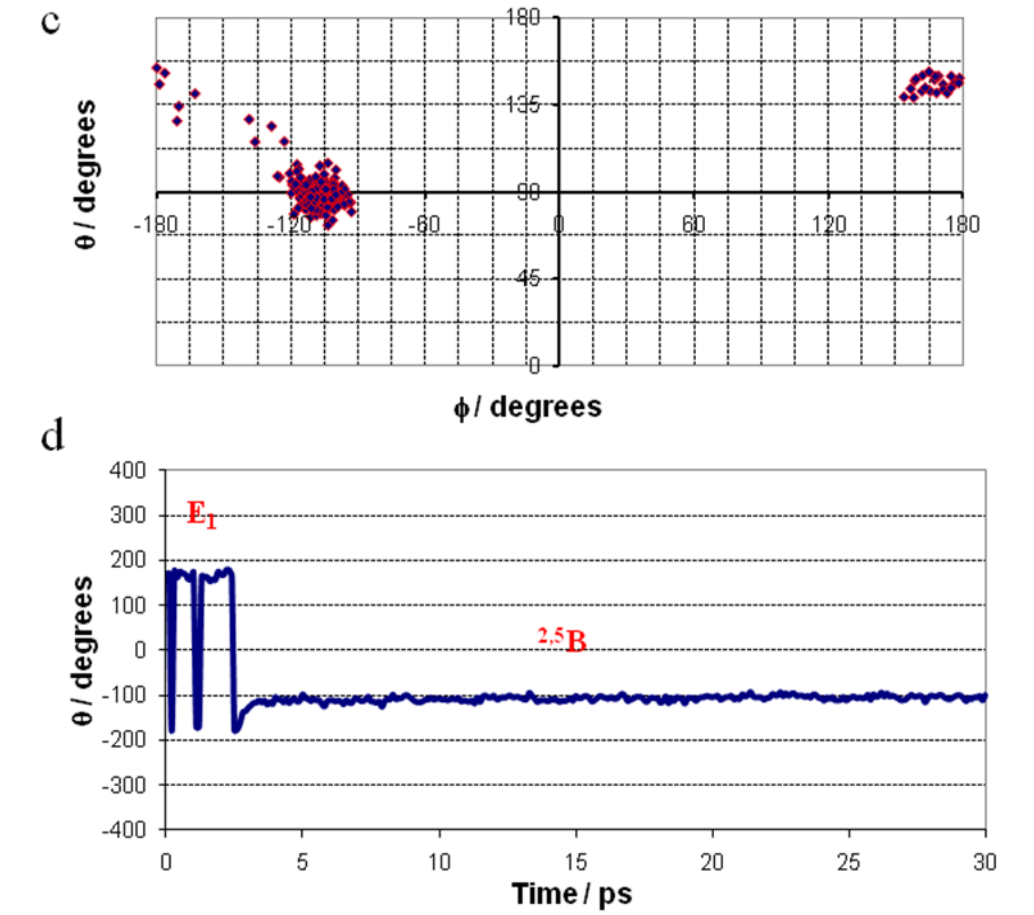


Fig. 5.9 QM/MM MD simulation for Model 3: (a) plot of composite angle ϕ *vs*. time and (b) conformational space sampled during the QM MD trajectory for xylose ring conformation starting from $^{2,5}B$ conformation; (c) plot of ϕ *vs*. time and (d) conformational space sampled starting from $^{4}C_1$ conformation.

Which residue(s) stands behind this conformational behavior?

The abrupt change in the conformational behavior from E_1 to $^{2,5}B$ does require an explanation, which we sought by consideration of four nearby residues that interact directly with the xylose ring at the −1 subsite, namely Glu172 and Glu78, Arg112 and Tyr69 (Fig. 5.10).

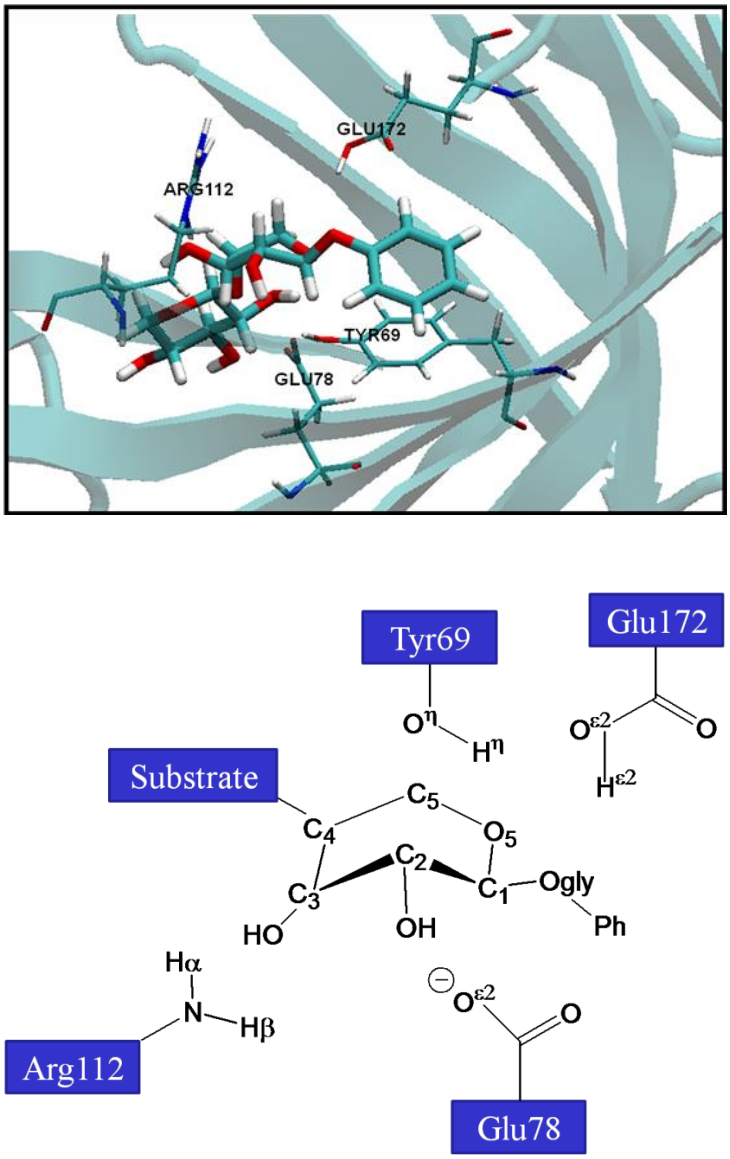
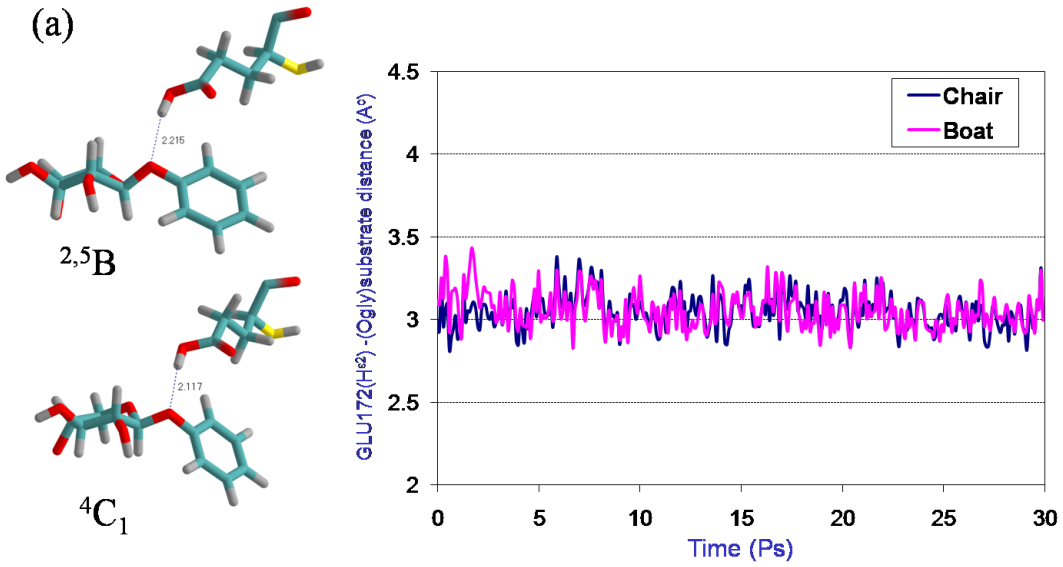
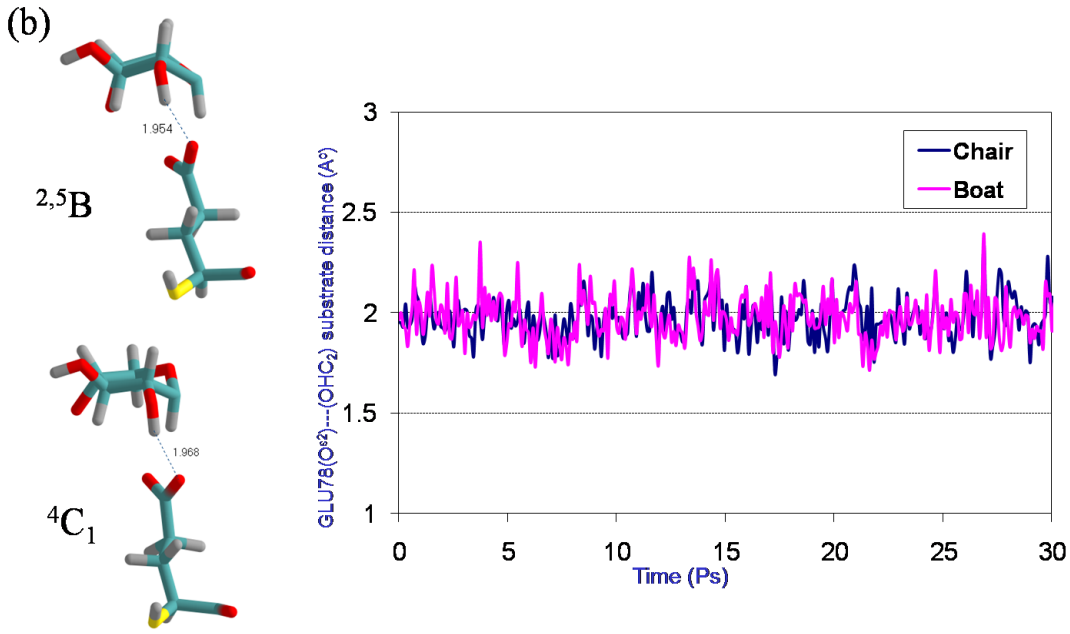


Fig. 5.10 Interaction of the proximal sugar ring with the active site residues

The acidic group Glu172 donates a hydrogen bond to the glycosidic O1 atom of the substrate (Fig.5.11a), whereas the nucleophilic group Glu78 accepts a hydrogen bond from the OH group at C2 of the sugar ring (Fig.5.11b), and Arg112 forms two hydrogen bonds with OH group at C3 of the substrate (Fig.5.11c). We have monitored these hydrogen-bond distances along the 30 ps MD trajectories for the chair and the boat conformations of the proximal sugar ring of the substrate, but have found essentially no difference between the two conformers. On the other hand, Tyr69 may form a hydrogen bond with either the endocyclic ring oxygen O5 or with Glu78 (Fig. 5.11d): when one hydrogen-bond

distance is short, the other is long (Fig. 5.12). In the QM/MM MD simulation started from the $^{4}C_{1}$ chair, the H1···O5 distance between Tyr69 and the substrate is initially > 3.5 Å and distinct from the short distance for the $^{2,5}B$ boat (Fig. 5.11d), but after 5 ps the two trajectories behave in the same way (Fig. 5.11d).





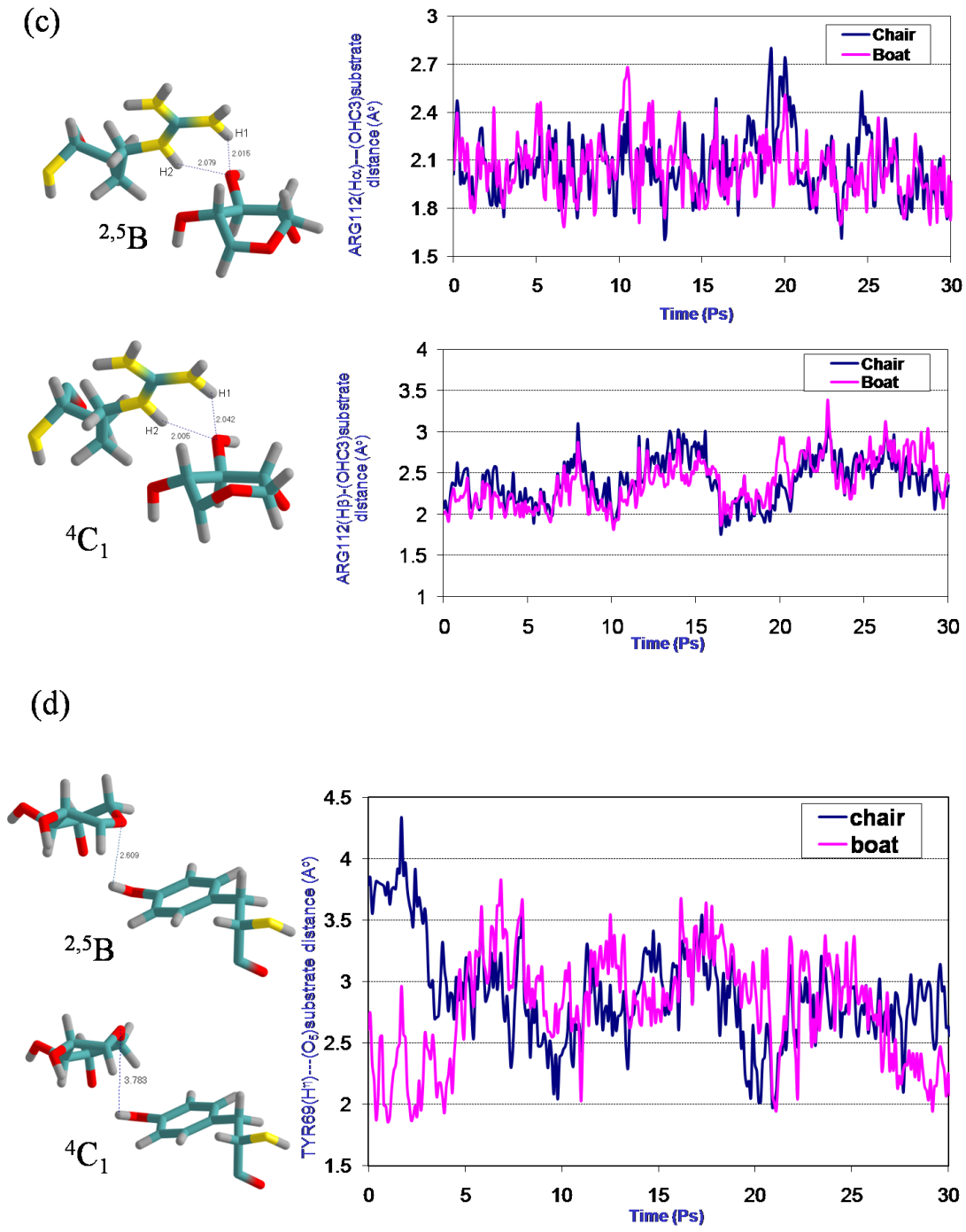


Figure 5.11: Analysis of hydrogen bond interaction between the sugar ring at subsite-1 and the active site residues accompanied with stereo diagram of each residue for model 3 $_{b \text{ and } c}$. From a to d, GLU172, GLU78, ARG112 and TYR69, respectively.

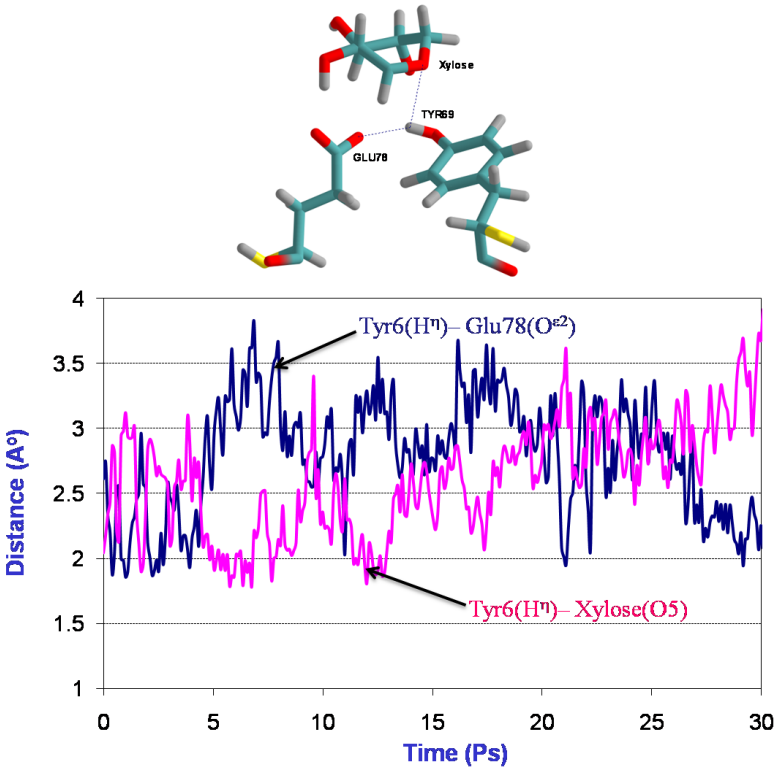


Fig.5.12 Tyr69 forms a hydrogen bond either to the endocyclic ring oxygen O5 of the xylose boat conformer (purple line) or to oxygen O_n of the Glu78 residue (blue) and (b) switches between the two during a QM/MM MD simulation.

The remarkable change in the hydrogen bond between Tyr69 and the proximal sugar ring suggests that this interaction is required to assist the conformational rearrangement from the $^4C_1$ chair to the $^{2,5}B$ boat in the enzyme active site. To test this hypothesis we made the Y69F mutant, replacing tyrosine by phenylalanine, in order to eliminate this hydrogen bonding interaction with the substrate.

Using the same computational procedure as before, QM/MM optimization showed the $^4C_1$ conformation to be a local minimum, and subsequent MD simulation confirmed that it remained as a stable species during the 30 ps trajectory (Fig. 5.13). This strongly indicates that Tyr plays a vital role in stabilizing the $^{2,5}B$ conformer.

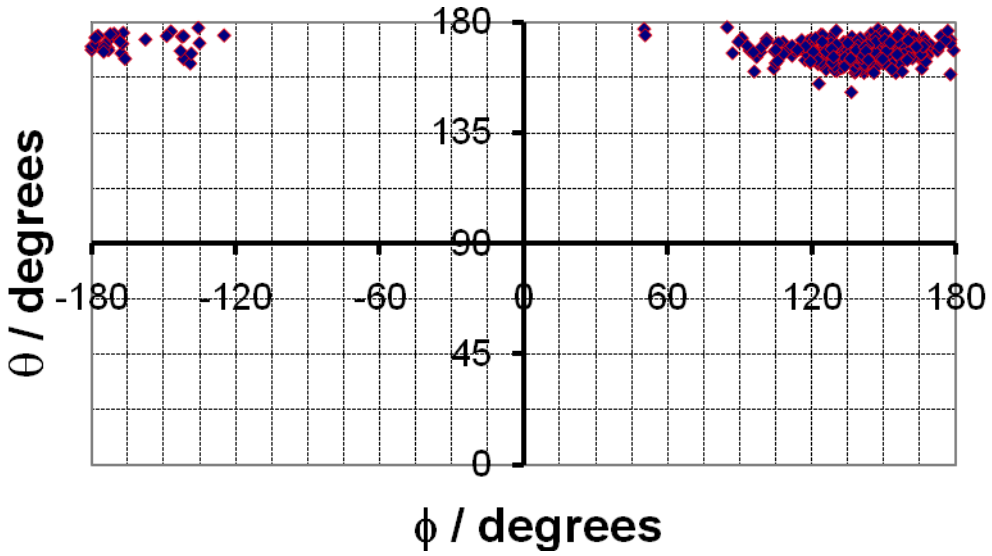


Fig. 5.13 Regions of xylose ring conformational space sampled during the MD trajectories for the $^{4}C_{1}$ of the Y69F mutant.

To quantify the extent of this stabilization, we computed free energy profiles (Fig.5.14) for the interconversion of the $^{2,5}B$ boat and the $^{4}C_{1}$ chair for both the wild-type enzyme and Y69F mutant enzyme-substrate complexes, using the C3-C4-C5-O5 dihedral as the reaction coordinate. The (Helmholtz) free changes along this coordinate are evaluated relative to the well-defined boat conformer with C3-C4-C5-O5 ≈ 45° for both profiles. Consequently, the energetic stabilization of the boat conformer in the wild-type due to the hydrogen bond between the substrate and Tyr69 appears in Fig. 5.14 as an apparent stabilization of the chair conformer in the mutant by ≈ 20 kJ mol^{-1}. Note that, for the wild-type, the dihedral angle C3-C4-C5-O5 ≈ –60° does not correspond purely to a $^{4}C_{1}$ chair: the population of conformers sampled in the simulation for values of the reaction coordinate in this range includes also envelope and half-chair conformations. It is therefore probably safer to estimate the degree of stabilization by comparison between the local energy minima at ≈ –60° for the mutant and ≈ –40° for the wild-type. Clearly, the barrier for conversion of the $^{4}C_{1}$ chair to the more-stable $^{2,5}B$ boat in the wild-type enzyme-substrate complex is significantly lower than it is for the Y69F mutant.

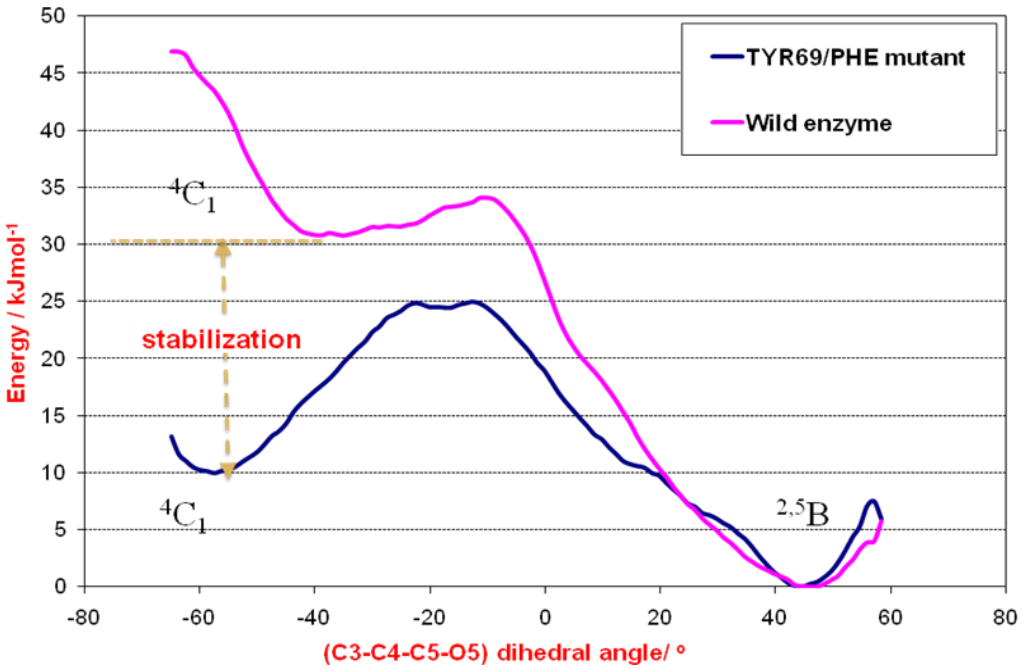


Fig. 5.14 Free energy profiles for interconversion of the $^{4}C_{1}$ chair and the $^{2,5}B$ boat conformers of the proximal xylose ring of the substrate in wild-type BCX and Y69F mutant .

The sequence of conformations populated by the wild-type BCX-substrate complex as it is driven along the C3-C4-C5-O5 dihedral angle reaction coordinate in the PMF calculations is presented in Fig. 2.15. The differently coloured clusters of points represent a selection of the overlapping windows considered during umbrella sampling; for the sake of clarity, not all the windows are plotted in this diagram. Starting from the $^{2,5}B$ boat, the sequence of MD trajectories sample conformations including $^{2}S_{O}$, $B_{O,3}$, E_{3}, $^{2}H_{3}$, ^{2}E, $^{2}H_{1}$, E_{1}, $^{O}H_{1}$, ^{O}E, $^{O}H_{5}$ and E_{3} on the way towards $^{4}C_{1}$. Note that this sequence of clusters does not define a dynamical trajectory since it is driven by the dihedral angle constraint and is not a function of time. Also note that the "polar" $^{4}C_{1}$ region is distorted by the Mercator projection in regard to area and distance relative to the other sugar-ring conformations.

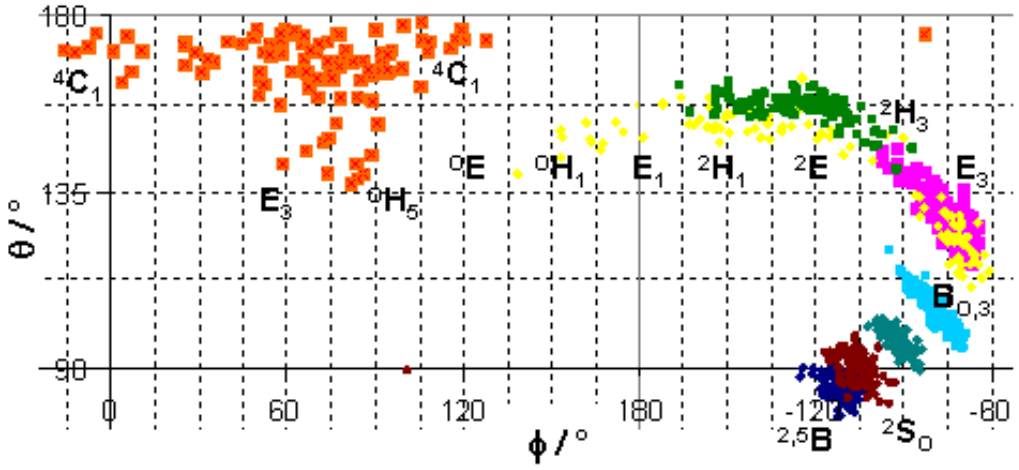


Fig. 5.15 Conformations sampled during the PMF calculation for ${}^{2,5}B \rightarrow {}^{4}C_1$ interconversion with respect to the C3-C4-C5-O5 dihedral angle in the wild-type enzyme-substrate complex.

Development of oxacarbenium-ion character

Mulliken atomic charges for selected atoms (Table 5.15) averaged over the 30 ps MD trajectory for each of the three models suggest a gradual increase in the oxacarbenium-ion character in the proximal xylose ring as the substrate interacts with the carboxyl/carboxylate pair in solution and within the wild-type enzyme active site. This is shown by increases in positive charge on C1 and H1 and a decrease of negative charge on O5; these changes are accompanied by increased polarization of the C1-O1 bond as shown by an increase of negative charge on O1. Note that these charge changes are more pronounced for the boat than for the chair conformer. Moreover this development of oxacarbenium-ion character is demonstrated by shortening and lengthening, respectively, of the average C1-O5 and C1-O1 bonds (Table 5.2); curiously, these changes are larger for the chair conformer.

The effect of the Y69F mutation is to reduce the degree of oxacarbenium-ion character in the proximal xylose ring of the enzyme-substrate complex, as evidenced by the charges and bond lengths shown in Tables 5.1 and 5.2. The non-bonded distances $O^{\varepsilon 2-}\cdots C1$ and $H^{\varepsilon 2}\cdots O1$ in Table 5.2 indicate that greater oxacarbenium-ion character is accompanied by a shorter distance to the nucleophilic carboxylate group but a longer distance to the acidic carboxyl group.

Table 5.1 Average atomic charges <*q*> / |*e*| on selected atoms for populations of boat and chair conformers.

	model 1		model 2		model 3		Y69F
atom	boat	chair	boat	chair	boat	chair	boat
C1	0.11	0.11	0.13	0.11	0.15	0.14	0.16
H1	0.15	0.13	0.18	0.14	0.23	0.21	0.17
O5	–0.28	–0.29	–0.24	–0.29	–0.22	–0.23	–0.27
O_{gly}	–0.25	–0.23	–0.27	–0.22	–0.28	–0.27	–0.26

Table 5.1 Average interatomic distances <*d*> / Å for populations of boat and chair conformers.

	model 1		model 2		model 3		Y69F
distance	boat	chair	boat	chair	boat	chair	boat
C1-O5	1.422	1.429	1.419	1.427	1.411	1.412	1.418
C1-O_{gly}	1.428	1.421	1.432	1.421	1.439	1.438	1.434
$O^{\varepsilon 2}$(Glu78)···C1			4.204	4.311	3.041	3.052	3.089
$H^{\varepsilon 2}$(Glu172)···O_{gly}			2.479	2.751	2.619	2.629	2.524
$O^{\varepsilon 2}$(Glu172)···$O^{\varepsilon 2}$(Glu78)			7.596	7.992	7.432	7.426	7.298

5.4. Discussion

The suggestion that a pyranoside substrate might preferentially adopt a $^{2,5}B$ conformation during enzyme-catalysed glycoside hydrolysis was first made by Hosie and Sinnott (*12*) on the basis of kinetic isotope effects for yeast α-glucosidase catalysed hydrolysis of aryl glucosides and glucosyl pyridinium ions. Their data required that breaking of the bond to the aglycone moiety of their substrates was preceded by a kinetically discrete non-covalent transformation of the initial enzyme-substrate complex, which they identified with a change to the $^{2,5}B$ conformation in which the C5-O5-C1-C2 atoms are approximately co-planar.

The latter arrangement would facilitate the formation of a transition state with significant oxacarbenium-ion character.

In this work we have taken one step toward investigating the reason for, as well as the structural implications of, substrate ring distortion in β-1,4-xylanase, by applying hybrid QM/MM molecular dynamics and free energy calculations. How reliable are our results, based as they are upon use of the semiempirical AM1 hamiltonian within the QM/MM method? Momany and co-workers (*13*) have shown that the B3LYP/6-311++G** level of theory gives consistently reliable geometries, conformations, and energies for carbohydrates *in vacuo*, and have noted the importance of hydroxyl group orientations and interactions. Intramolecular interactions between these groups are less important in aqueous solution because of hydrogen bonding with solvent molecules, and in an enzyme active site interactions with the protein environment may drastically reduce the number of significant hydroxyl group rotameric forms. In regard to the relative energies of the $^{4}C_{1}$ and $^{2,5}B$ conformations of the xylose ring in each of our condensed phase models 1, 2 and 3, one may well suppose that there is an underlying systematic error due to AM1 as compared to a large-basis DFT, or other high-level theoretical method, but it is not a straightforward task to estimate its magnitude. Moreover, our main concern in this study is not with the energy difference between the boat and chair conformers, but rather with the influence of enzyme environment upon that energy difference. We expect that the change from tyrosine to phenylalanine (both in the MM region) will polarize higher-level QM wavefunctions in a qualitatively similar (though undoubtedly quantitatvely different) fashion to AM1. We therefore consider our AM1/OPLS estimate of the free energy of stabilization of the boat conformer due to the hydrogen bond with Tyr69 in BCX to be at least qualitatively reliable. A quantitative assessment of error, as compared to an appropriate high-level QM method, would be meaningless unless it were evaluated for completely converged populations of conformations accessible under condensed-phase conditions, since there are large fluctuations between energy differences determined at arbitrary "snapshot" structures taken from the MD trajectories.

The first structural evidence for sugar-ring distortion to the boat conformer was obtained by Brayer and co-workers (*1*) who obtained the 1.8 Å resolution

structure of the covalent glycosyl-enzyme intermediate formed between BCX and a 2-deoxy-2-fluoro-β-xylobioside. (This structure was, of course, the starting point for the modelling studies described in the present work.) The proximal xylose ring was found to be heavily distorted from $^{4}C_{1}$ to $^{2,5}B$, thus achieving near coplanarity of the C5-O5-C1-C2 atoms, whereas the distal xylose ring remained in the chair conformation. It was noted that an itinerary for the conformational rearrangement that were to proceed via $^{4}C_{1} \rightarrow {}^{2}H_{3} \rightarrow {}^{2}S_{O} \rightarrow {}^{2,5}B$ would involve motion (relatively) of only C5 and O5, and that these atoms do not bear substituents which might add to the energetic cost of the conformational change. These workers also obtained the X-ray crystal structure for the catalytically inactive Y69F mutant BCX, with no substrate in the active site, and reported that it has a very similar structure to that of the covalent intermediate. In particular, the position of Glu78 was almost identical, thus refuting an earlier proposal (*14*) that Tyr69 might have a crucial catalytic role in correctly positioning the nucleophile. On the other hand, it was noted (*1*) that the non-covalent enzyme-substrate complex observed by Wakarchuk *et al .(14)* between the catalytically incompetent E172C mutant BCX and a xylotetraose substrate has the proximal xylose ring (occupying the −1 subsite) bound in the $^{4}C_{1}$ conformation.

What none of these prior experimental studies has been able to ascertain is the nature of the conformation of a xylose substrate bound non-covalently in the −1 subsite of a catalytically competent BCX, but that is precisely what we have been able to achieve by means of the computational modelling carried out in the present study. Our QM/MM model for the enzyme-substrate complex between wild-type BCX and phenyl β-xylobioside clearly prefers the proximal ring to adopt the $^{2,5}B$ conformation. An MD trajectory initiated from a structure with this ring in the $^{4}C_{1}$ conformation spontaneously transformed into the $^{2,5}B$ boat. The PMF for interconversion of these conformers along the dihedral angle C3-C4-C5-O5 coordinate suggests that the boat is lower in free energy than the chair site by ~30 kJ mol^{-1} within the active site of wild-type BCX.

We have demonstrated the key role of hydrogen bond donation from Tyr69 to O5 of the proximal xylose ring in stabilizing the boat over the chair by means of a comparative QM/MM MD study for the Y69F mutant BCX which clearly shows

the $^4C_1$ conformation to be a stable local energy minimum in the absence of that hydrogen bond. The PMF for interconversion of boat and chair along the dihedral angle C3-C4-C5-O5 coordinate suggests that the boat is still lower in free energy than the chair site by ~10 kJ mol^{-1} within the active site of the Y69F mutant, implying a preferential stabilization of the boat by wild-type BCX by about 20 kJ mol^{-1}.

The $^{2,5}B$ conformation of phenyl β-xyloside is not stable in the gas phase but seems to be metastable in water, being able to fluctuate within at least the $^{2,5}B \rightarrow {}^2S_0 \rightarrow B_{3,0}$ section of the boat/skew-boat pseudorotational itinerary. In the presence of an added carboxylate/carboxyl pair, the boat is a local minimum in the gas-phase and in water it fluctuates between $^{2,5}B$ and $^2S_0$: a hydrogen bond between the OH group at C2 of xylose and the propionate moiety seems to provide some stabilization for the $^{2,5}B$ conformation. As expected, the $^4C_1$ chair is stable in water both with and without the propionate/propionic acid pair. The Y69F mutant has been reported (*14*) to have less than 0.01% of the activity of wild-type BCX at 40 °C, which would imply a free energy difference between the rate-determining transition states of > 24 kJ mol^{-1}. It has been hypothesized (*1, 15*) that the catalytic role of Tyr69 is provide a stabilizing dipolar interaction with the partial positive charge on O5 of the oxacarbenium-ion like transition state for the rate-determining glycosylation of BCX by the substrate. It is worth noting an alternative possibility that arises by analogy with the experimental (*16*) and theoretical (*17*) studies by Schramm, Schwartz and co-workers on human purine nucleoside phosphorylase (hPNP). The glycosyl transfer reaction catalysed by this enzyme has a transition state with oxacarbenium-ion character and the ribofuranoside substrate possesses a hydroxyl group C5. It is suggested that the neighbouring His257 provides a mechanical push upon O5 towards the endocyclic O4 in a compressive motion with the phosphate nucleophile such that the build-up of electron density stabilizes the oxacarbenium-like transition state and facilitates the reaction (*16, 17*). Regardless of whether this particular idea is correct, it is striking to note the importance of a suitably located hydroxyl group which exerts on effect upon the endocyclic oxygen of a glycoside by means of its oxygen rather than its proton. Of course, the xylose ring in the −1 subsite of BCX has no hydroxymethyl substituent at C5; indeed, it has been noted that there is no

space around C5 to accommodate any substituent (*1*). Instead, however, the active site of wild-type BCX presents Tyr69 in close proximity to O5. We suggest that the OH groups of Tyr69 in BCX and of the hydroxylmethyl substituent of the ribofuranoside substrate of hPNP may have similar roles, this will be discussed in details in chapter 7.

5.5. Conclusion

Molecular dynamics simulations using a hybrid QM/MM potential have demonstrated that wild-type BCX preferentially binds a phenyl β-xylobioside substrate in the $^{2,5}B$ conformation at the −1 subsite adjacent to the scissile glycosidic bond. This result complements and extends the earlier experimental report that the proximal xylose ring of a covalently-bound 2-deoxy-2-fluoroxylobioside complex with BCX adopts the same $^{2,5}B$ boat conformation. Analogous simulations for the Y69F mutant confirm the key role of Tyr69 in stabilizing the boat in preference to the $^{4}C_{1}$ chair conformation with a relative free energy difference of about 20 kJ mol^{-1}.

References

1. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase, *Biochemistry*, **38**, 5346-5354.
2. Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005) H++: a server for estimating pK(a)s and adding missing hydrogens to macromolecules, *Nucleic Acids Res.*, **33**, W368-W371.
3. Field, M. J., Albe, M., Bret, C., Proust-De Martin, F., and Thomas, A. (2000) The Dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials, *J. Comput. Chem.*, **21**, 1088-1100.
4. Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. (1985) The development and use of quantum-mechanical molecular models .76. AM1 - a new general-purpose quantum-mechanical molecular model, *J. Am. Chem. Soc.*, **107**, 3902-3909.
5. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, *J. Phys. Chem. B*, **105**, 6474-6487.
6. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2000) OPLS-AA/L force field for proteins: Using accurate quantum mechanical data., *Abstr. Papers Am. Chem. Soc.*, **220**, U279-U279.
7. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, **79**, 926-935.
8. Cremer, D., and Pople, J. A. (1975) General definition of ring puckering coordinates, *J. Am. Chem. Soc.*, **97**, 1354-1358.
9. Biarnes, X., Ardevol, A., Planas, A., Rovira, C., Laio, A., and Parrinello, M. (2007) The conformational free energy landscape of beta-D-glucopyranose. implications for substrate preactivation in beta-glucoside hydrolases, *J. Am. Chem. Soc.*, **129**, 10686-10693.

10. Berces, A., Whitfield, D. M., and Nukada, T. (2001) Quantitative description of six-membered ring conformations following the IUPAC conformational nomenclature, *Tetrahedron*, **57**, 477-491.

11. Gaussian 03, R. C., M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004.

12. Hosie, L., and Sinnott, M. L. (1985) Effects of deuterium substitution alpha and beta to the reaction center, O-18 substitution in the leaving group, and aglycone acidity on hydrolyses of aryl glucosides and glucosyl pyridinium ions by yeast alpha-glucosidase - a probable failure of the antiperiplanar-lone-pair hypothesis in glycosidase catalysis, *Biochemical J.*, **226**, 437-446.

13. Schnupf, U., Willett, J. L., Bosma, W. B., and Momany, F. A. (2007) DFT study of alpha- and beta-D-allopyranose at the B3LYP/6-311++G** level of theory, *Carbohydrate Res.*, **342**, 196-216.

14. Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J., and Yaguchi, M. (1994) Mutational and crystallographic analyses of the active-site residues of the *Bacillus circulans* xylanase, *Protein Sci.*, **3**, 467-475.

15. Zechel, D. L., and Withers, S. G. (2000) Glycosidase mechanisms: Anatomy of a finely tuned catalyst, *Acc. Chem. Res.*, **33**, 11-18.

16. Murkin, A. S., Birck, M. R., Rinaldo-Matthis, A., Shi, W. X., Taylor, E. A., Almo, S. C., and Schramm, V. L. (2007) Neighboring group participation in the transition state of human purine nucleoside phosphorylase, *Biochemistry*, **46**, 5038-5049.

17. Nunez, S., Antoniou, D., Schramm, V. L., and Schwartz, S. D. (2004) Promoting vibrations in human purine nucleoside phosphorylase. A molecular dynamics and hybrid quantum mechanical/molecular mechanical study, *J. Am. Chem. Soc.*, **126**, 15720-15729.

6. Exploring the catalytic mechanism

Overview

In this chapter we perform various computational approaches using hybrid quantum-mechanical/molecular-mechanical potentials to investigate the catalytic mechanism of the retaining endo-β-1, 4-xylanase (BCX) from *B. circulans*.

Chapter 3 provides a more detailed background about the proposed catalytic pathway and the experimental studies for endo-β-1, 4-xylanase (BCX) from *B. circulans.* Figure 6.1 reviews the first step in the catalytic course, glycosylation, as we focus upon this part in our work.

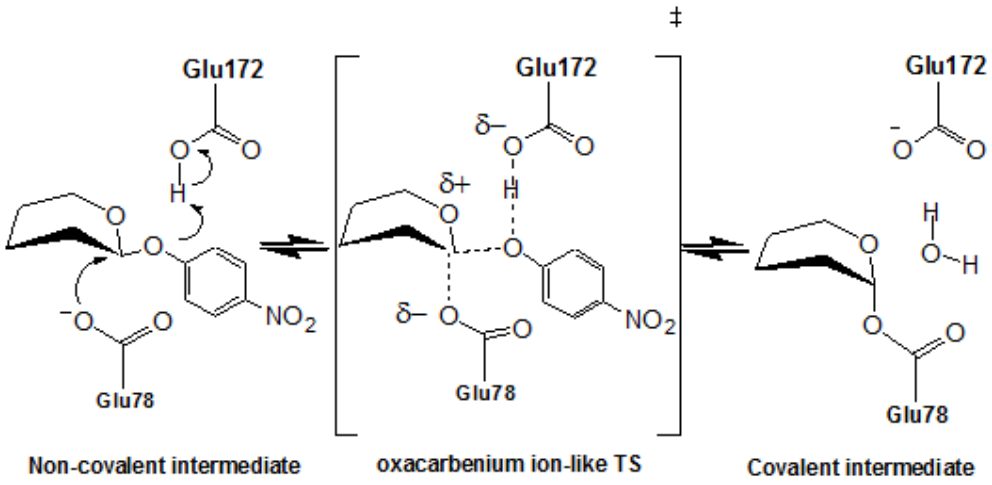


Fig.6.1. Formation of glycosyl-enzyme intermediate.

At the end of this chapter, we also report some preliminary kinetic isotopic effect calculations for BCX.

6.1. Modelling of Reaction Mechanism

6.1.1. Computational Methods

Model System

As mentioned in chapter 5, we initiate the modelling from the PDB structure for the covalent enzyme-inhibitor complex (Figure 6.1) of wild-type BCX (accession code 1BVV) as this fulfils several criteria: it is a high-resolution (1.8 Å) crystal

structure; the sugar substrate is covalently bound to the enzyme, thus allowing us to avoid errors involved with docking the substrate into the active site; the stereochemical configuration of the substrate is clearly defined; there are no mutations within the protein. A non-covalent reactant complex (RC) with a xylobioside with a *p*-nitrophenyl (PNP) leaving group attached to anomeric carbon (C1, but hereinafter denoted C_{ano}) of the proximal sugar was constructed as previously described in chapter 5. We are not concerned here with qualitative and quantitative influences of leaving group pK_a variation on the 2-D potential/free-energy landscape; these issues will be discussed in conjunction with kinetic isotope calculations in the near future.

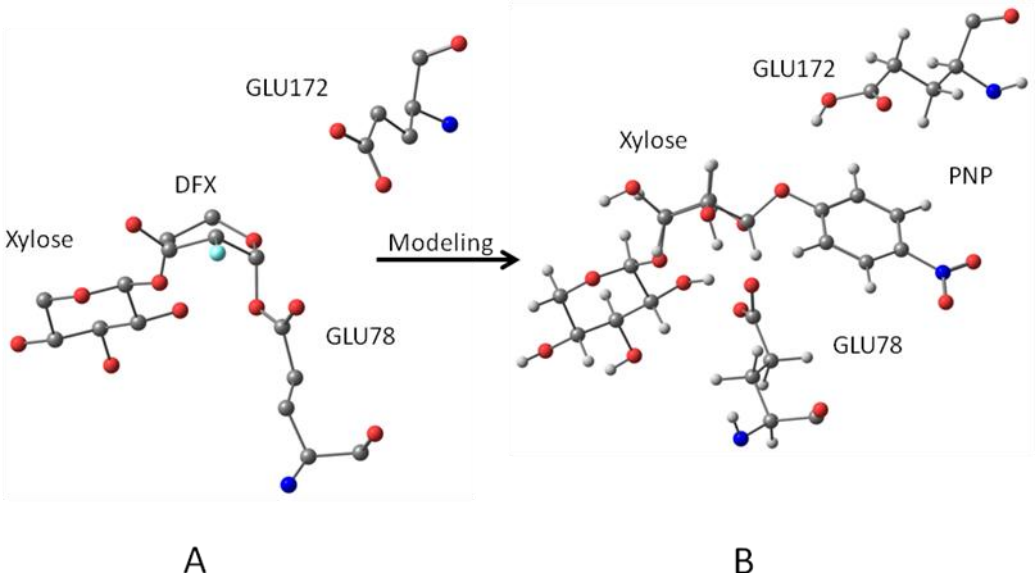


Fig.6.2. BCX active site from 1BVV crystal structure of covalent complex (A) and as modified for the non-covalent reactant complex (B).

QM/MM Simulations

All simulations were done using the Dynamo library (*1*) of programs for molecular dynamics (MD). Except for transition-structure (TS) localization, hessian calculation, and normal mode analysis (see below), the enzyme system was divided into two regions (Figure 6.3) in all simulations: the QM region contained active-site atoms described by the AM1 semi-empirical Hamiltonian (*2*), and the MM region contained atoms in the rest of the enzyme and the surrounding solvent molecules described by the OPLS-AA potential (*3*). For most simulations, the QM region contained 70 atoms consisting of the substrate,

a xylobioside (XYL) with PNP aglycon, and the two catalytic residues Glu172 and Glu78. In some simulations we included the Tyr69 residue in the QM region; this larger model had QM 86 atoms. All covalent bonds between atoms of the QM and MM regions were treated by the link-atom approach in the Dynamo program: QM link atoms were placed along the C_β(QM)–C_α(MM) bonds of Glu78 and Glu172 and, where appropriate, Tyr69. The whole system was enveloped in a cubic box of TIP3P water (*4*) of side-length 55.5 Å. The total simulation system had 16476 atoms. QM/MM calculations were performed to obtain 2-D free energy surfaces, to locate saddle-points, to determine reaction paths, and to evaluate Hessians.

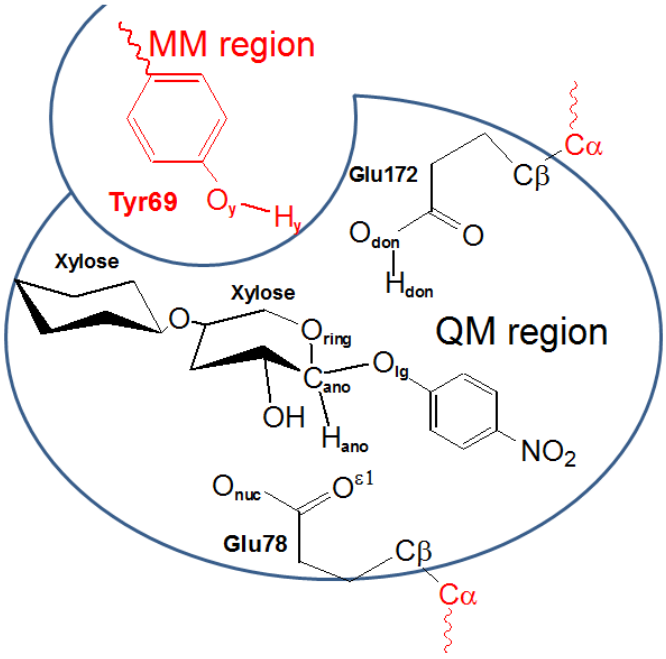


Fig.6.3. QM/MM partitioning.

2D-PMF Calculations

To construct a two-dimensional free energy surface for the glycosylation step, potentials of mean force (PMFs) were computed using the method of umbrella sampling (*5*) as follows. Two composite reaction coordinates, ξ_1 and ξ_2, which

respectively describe nucleophilic substitution and proton transfer, were found (after many trials with other geometrical parameters) to give a good description of the mechanistic process; the oxygen of glycosidic bond to the leaving is denoted as O_{lg} whereas $O^{\varepsilon 2}$ of the Glu172 proton donor is denoted as O_{don}. It was important to consider both variables in order to allow for possible coupling of proton transfer with nucleophilic attack. An umbrella potential U_i of harmonic form (eq 3) was defined for each coordinate ξ_i, where k_i is a force constant and ξ_i^0 is a constant reference value.

$$\xi_1 = d(O_{lg}\cdots C_{ano}) - d(O_{nuc}\cdots C_{ano}) \quad (1)$$

$$\xi_2 = d(O_{don}\cdots H_{don}) - d(H_{don}\cdots O_{lg}] \quad (2)$$

$$U_i(\xi_i) = \tfrac{1}{2}k_i(\xi_i - \xi_i^0)^2 \quad (3)$$

All MD simulations employed the QM/MM potential described in the previous section and were performed in a series of windows using $k_1 = k_2 = 2500$ kJ mol^{-1} Å^{-2}, and with ξ_1^0 and ξ_2^0 being varied in increments of 0.1 Å from –1.64 to +3.0 Å and from –1.55 to +1.0 Å, respectively. The simulation in each window was ~10 ps length, with 2 ps equilibration: these values were found adequate to ensure sufficient overlapping of the simulations in neighboring windows along each reaction coordinate in order to obtain satisfactory convergence of the PMF. All simulations were performed in the *NVT* ensemble at a temperature of 300 K; a velocity-Verlet-Langevin algorithm was used with a friction coeffcient of 10 ps^{-1} for each atom and a time step of 1 fs (*6*). The nonbonding interactions were calculated using an atom-based force-switching truncation function with inner and outer cutoffs of 10.5 and 11.5 Å, respectively. At each step of the data collection, the values of ξ_1 and ξ_2 were saved on an external file. The reaction-coordinate distribution functions were then determined for each window and pieced together using the weighted histogram analysis method (*7*). The resulting distribution function $\rho(\xi_1, \xi_2)$ was then used to calculate the PMF, W (eq 4), where k_B is the Boltzmann constant, T is the temperature, and c is an arbitrary additive constant (*8*). Differences in W provide estimates for changes in the

Helmholtz free energy which, for a condensed phase system, are essentially equivalent to Gibbs energy changes.

$$W(\xi_1, \xi_2) = c - k_B T \ln \langle \rho(\xi_1, \xi_2) \rangle \quad (4)$$

TS and Reaction-Path Calculations

Since it was not feasible to perform free-energy calculations directly with density-functional theory (DFT) techniques for a system of the size considered here, instead we estimated the probable errors in the QM part of the QM/MM calculations by means of the following procedure.

A 1-D AM1/OPLS energy profile was generated in turn for each step of the glycosylation in the full system of enzyme, substrate and solvent. Up to 15 structures in the vicinity of the relevant transition state on the AM1/OPLS 2D-free-energy surface were selected, each of which was refined to a first-order saddle point (a TS) using standard second-derivative saddle-point location techniques. The QM region together with 17 active-site residues of the immediately surrounding MM region were optimized using second-derivative techniques, and the remainder of the system was optimized using a standard gradient minimizer, as implemented in Dynamo. Normal-mode analyses were done on all stationary points, and visual inspection of the normal mode displacements was carried out with the Molden program (*9*). All starting points were found to converge to one and the same TS. From this refined TS a sequence of structures along the chosen reaction coordinate (in turn either ξ_1 or ξ_2) was generated by incrementing and decrementing the reference value ξ_i^0 by 0.05 Å and, at each point, reminimizing the whole system subject to a very strong harmonic constraint (k_i = 20000 kJ mol^{-1} Å^{-2}) applied to that reaction coordinate variables close to its reference value while allowing the rest of the system to minimize freely. This process allowed a continuous reaction path (similar to, but not identical with, an intrinsic reaction coordinate) to be generated between initial and final structures for each step of the reaction. Owing to the nature of this procedure note, however, that the final structure for the nucleophilic substitution was not exactly the same as the initial structure for the proton transfer.

To obtain higher-level corrected values for the reaction barriers, single-point energies were calculated for (only) the QM atoms of the reactant, TS, and product structures in vacuum with AM1 and with a DFT method; a total of 11 structures was adequate to span the complete range from initial to final for each step. The DFT calculations were performed by means of the Gaussian03 program (*10*) with B3LYP (*11*) and the 6-31+G* basis set. Fully optimized TS structures and reaction paths were computed in vacuum for a truncated model of 53 QM atoms comprising the proximal sugar ring and leaving group of the substrate, together with propionate and propionic acid moieties to represent the Glu78 and Glu172 residues of the active site.

6.1.2. Results and discussion

Reaction Mechanism for Wild-Type BCX

Figure 6.4 shows the 2-D surface of free-energy, with respect to the composite reaction coordinates ξ_1 for nucleophilic substitution at C_{ano} and ξ_2 for proton transfer to O_{lg}, for glycosylation of wild-type BCX by a PNP xylobioside substrate; this was generated from 846 individual AM1/OPLS MD trajectories requiring >10 ns of simulation in total. It was important to consider both variables in order to allow for possible coupling of proton transfer with nucleophilic attack in a concerted, general-acid catalyzed mechanism. However, the preferred pathway, shown by the dashed line on the Figure, proceeds in a stepwise fashion from RC with nucleophilic substitution occurring first, with a barrier of 176 kJ mol^{-1}, to form the glycosyl-enzyme intermediate INT via TS1. The second step, from INT to PC via TS2, involves the proton transfer from Glu172 to the O_{lg} of the PNP anion with a barrier of 79 kJ mol^{-1}.

Withers and coworkers (*13, 14*) reported linear Bronsted correlations of log k_{cat}/K_m with aglycone pK_a for wild-type BCX and for Glu78Cys carboxymethylated at Cys78, Glu172Asp and Glu172Cys carboxymethylated at Cys172 mutants with slopes of –0.7 or –0.8 for hydrolysis of aryl xylobiosides with good leaving groups. Their interpretation was that there was very substantial glycosidic bond cleavage at the glycosylation transition state with very little proton donation; the Glu172Cys and Glu172Gln mutants lacking a proton-donor

group were still capable of hydrolyzing substrates with good leaving groups (pK_a < 5.5) although they had no measurable activity with phenyl xylobioside or the natural substrate xylan. In the light of these experimental kinetic studies, our computational result for PNP xylobioside seems entirely reasonable. However, the calculated AM1/OPLS free-energy barrier is much too high: the experimental value for k_{cat} = 24 s^{-1} for reaction of wild-type BCX with this substrate at 25 °C (*12*) corresponds to $\Delta G^{\ddagger} \approx 65$ kJ mol^{-1}.

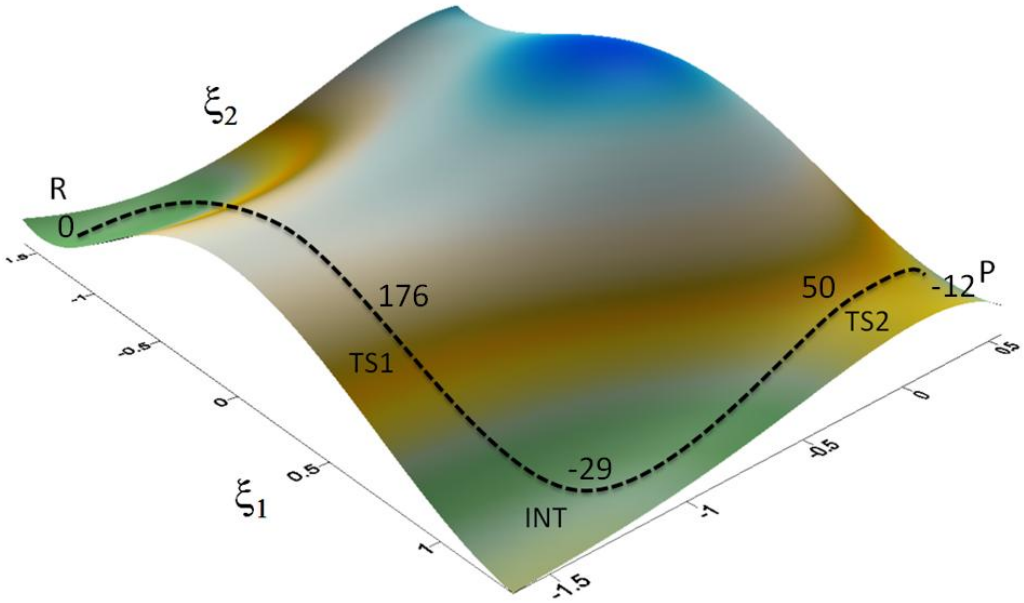


Fig.6.4. AM1/OPLS calculated 2D-free-energy profile for glycosylation as a function of the coordinates for nucleophilic substitution ξ_1 and proton transfer ξ_2 (relative energies in kJ mol^{-1}).

TS and Reaction-Path Calculations

Figure 6.5A shows the AM1/OPLS energy profile (solid line) along the reaction coordinate ξ_1 for nucleophilic substitution in the full model including the MM environment of the protein and solvent. The enthalpic barrier height $\Delta H^{\ddagger}$ = 150

kJ mol^{-1} from this 1-D calculation is somewhat lower than the free-energy barrier height $\Delta G^{\ddagger}$ = 176 kJ mol^{-1} from the 2-D calculation; the difference suggests a small decrease in entropy in the transition state. The dashed line shows the corresponding gas-phase AM1 energy profile for the QM atoms only with geometries frozen as in the full enzymic model; the curve shown is a fit to a series of 11 single-point energy calculations for structures along the reaction coordinate on either side of TS1. The barrier height is a little lower than for the full model, but the overall enthalpy change from RC to INT underestimates the exothermicity; this highlights the importance in QM/MM treatments not only of E_{QM} (the energy of the QM region) but also of E_{MM} and $E_{QM/MM}$ (the energy of the MM region and the interaction energy between the two regions).

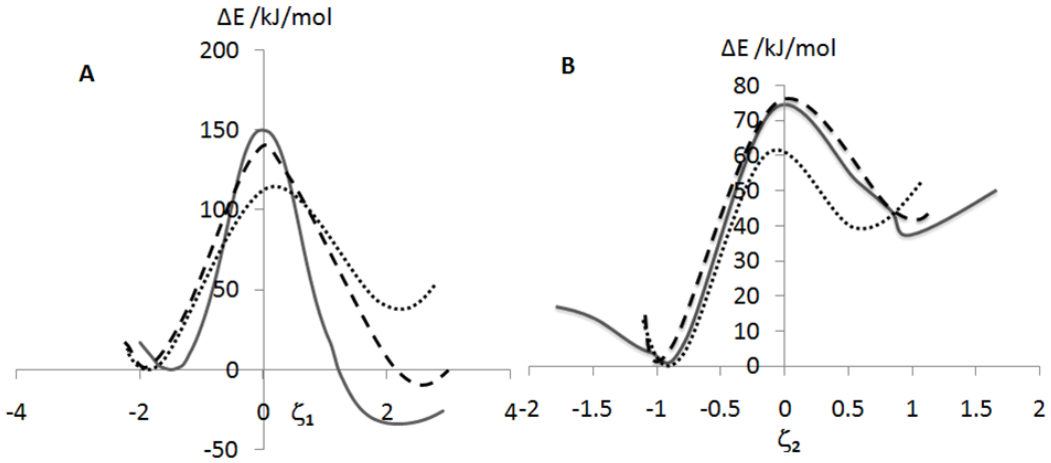


Fig.6.5. Relative energies for nucleophilic substitution (A, with respect to RC) and proton transfer (B, with respect to INT) calculated using AM1/OPLS in enzyme (solid line), AM1 single points (dashed line), and B3LYP/6-31+G* single points (dotted line).

The dotted line in Figure 6.5A shows the gas-phase B3LYP/6-31+G* energy profile for the QM atoms only, fit to single-point energies for the same 11 frozen geometries as above. Although the DFT potential energy barrier is appreciably lower than the AM1 barrier, the overall reaction energy change is predicted to be considerably endothermic; this demonstrates the inadequacy of using a "high-level" DFT method to estimate a correction to the "low-level" semi-empirical

QM/MM result if only E_{QM} is considered and if geometrical relaxation is ignored.

Figure 6.5B shows energy profiles along the reaction coordinate ξ_2 for proton transfer from Glu172 to O_{lg} in the full AM1/OPLS model (solid line) and for the gas-phase single-point energies calculated for the QM atoms only with AM1 (dashed line) and B3LYP/6-31+G* (dotted line). In this case all three methods give similar barrier heights and overall reaction energies. The profiles for nucleophilic substitution (Figure 6.5A) cannot be combined simply with those for proton transfer (Figure 6.5B) since the product geometry INT for the former is not quite the same as the reactant geometry INT for the latter, owing to the way that the profiles were computed with applied constraints.

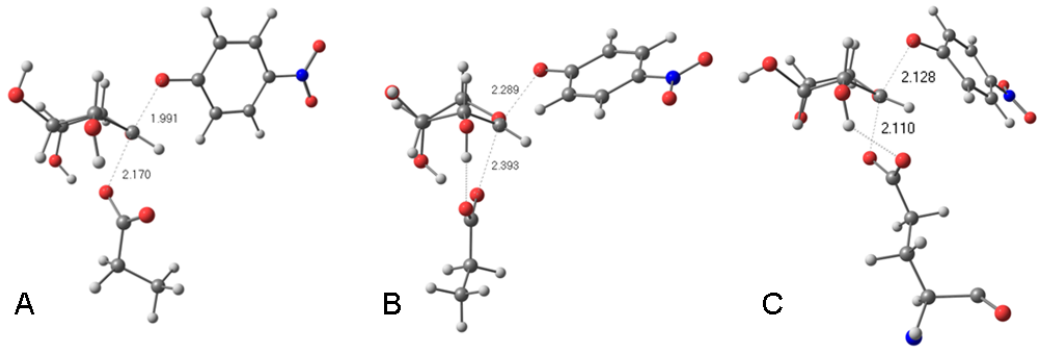


Fig.6.6. Transition structures for the nucleophilic substitution optimized using three different methods: gas-phase AM1 (A), gas-phase B3LYP/6-31+G* (B), and enzymic AM1/OPLS (C).

It is of interest to consider the AM1 or B3LYP/6-31+G* geometries and relative energies of RC, TS1 and INT optimized in vacuum without any constraints for the nucleophilic substitution and proton transfer steps. To this end we performed unconstrained TS optimizations with both methods for a model system of 53 QM atoms comprising the proximal sugar ring and leaving group of the substrate, together with propionate and propionic acid moieties to represent the Glu78 and Glu172 residues of the active site. Figure 6.6 and Table 6.1 show the resulting

geometries of TS1 for the nucleophilic substitution step for both methods alongside that for the full QM/MM model. It is remarkable that they are all very similar. The gas-phase AM1 TS geometry (Figure 6.6A) has shorter distances for the breaking O_{lg}····C_{ano} and making O_{nuc}····C_{ano} bonds than does the AM1/OPLS TS geometry (Figure 6.6C), whereas the B3LYP/6-31+G* TS geometry (Figure 6.6B) has longer distances. However, the nucleophilic carboxylate and acidic carboxyl moieties are located in essentially the same positions in all three TSs: this suggests that in the enzyme the catalytic residues are positioned more or less exactly where they prefer to be in vacuum. Furthermore, there is a hydrogen bonding interaction between $O^{\epsilon 1}$ of the nucleophilic carboxylate and the hydroxyl group at C2 of the proximal xylose ring: kinetic analysis indicates that this hydrogen bond contributes at least 18 kJ mol^{-1} to stabilization of the transition state for glycosylation in the retaining β-glucosidase from *Agrobacterium faecalis* (*15*). Inspection of the interatomic distances presented in Table 6.1 suggests that the key geometrical features of the substrate and its interactions with catalytic groups within the enzyme active site are well described by the AM1/OPLS method, as compared with freely optimized structures obtained in vacuum using both the AM1 and B3LYP/6-31+G* methods.

Table 6.1. Selected interatomic distances (Å) and transition frequencies for the QM region in different simulation methods. RC, TS1, INT, TS2, and PC refer to the reactant complex, transition structure for the nucleophilic attack, xylosyl-intermediate, transition structure for proton transfer, and product complex, respectively.

	AM1/OPLS (enzyme)					AM1 (gas-phase)					B3LYP/6-31+G* (gas-phase)				
	RC	TS1	INT	TS2	PC	RC	TS1	INT	TS2	PC	RC	TS1	INT	TS2	PC
C_{ano}···O_{lg}	1.44	2.13	3.71	4.61	4.30	1.44	1.99	2.96	3.01	4.90	1.47	2.28	4.11	4.22	4.31
O_{nuc}···C_{ano}	2.97	2.11	1.44	1.44	1.44	2.98	2.17	1.44	1.44	1.44	3.16	2.39	1.46	1.44	1.44
H_{don}···O_{lg}	3.21	2.95	3.95	1.15	0.98	2.08	1.98	1.91	1.13	0.99	1.85	1.58	1.57	1.20	0.98
O_{nuc}···O_{don}	6.96	7.63	7.93	8.31	8.13	6.76	6.5	6.44	6.71	7.20	6.64	6.11	6.34	6.45	6.57
$\nu^{\ddagger}/cm^{-1}$		460*i*		953*i*			413*i*		744*i*			119*i*		452*i*	

Another point of interest is the distance $O_{nuc}\cdots O_{don}$ between the nucleophilic carboxylate and acidic carboxyl groups, corresponding to Glu78 and Glu172 in BCX. As Table 6.1 shows, in the AM1/OPLS optimized structures for RC, TS1 and INT of the nucleophilic substitution step this separation is found to vary between about 6 and 7 Å. The AM1 and DFT structures optimized in vacuum yield separations between 6.4 and 6.8 Å and 6.1 and 6.6 Å, respectively. It should be noted that, while all of these distances are greater than the often-quoted value of 5.5 Å for retaining β-glycosidases (*16*), the structures considered here all contain the PNP aglycon group whereas this not present in any of the X-ray crystallographic structures used to evaluate the experimental average separation between the catalytic groups.

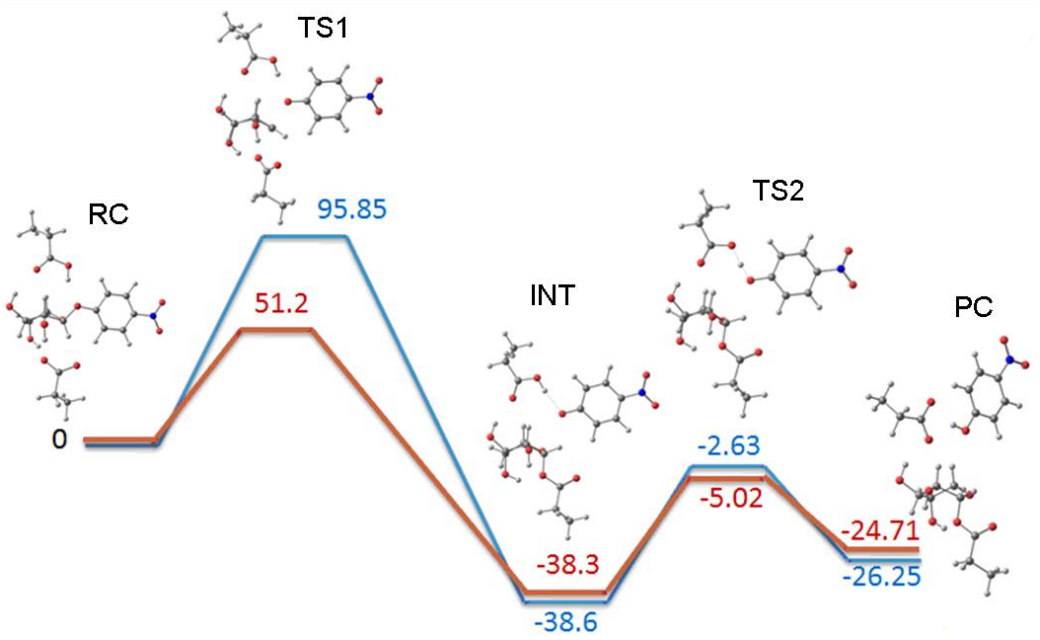


Fig.6.7. Energy profile for a simplified model of glycosylation using B3LYP/6-31+G* (red) and AM1 (blue) methods in vacuum.

Figure 6.7 shows the corresponding energy profiles in vacuum. In these cases the intermediate is common to both parts of the profile, since the no constraints were applied on these calculations. The overall energy change for formation of the covalent intermediate in this simple model is essentially the same for both

methods, and very similar to the QM/MM energy change for the full enzyme. However, the DFT potential energy barrier $\Delta E^{\ddagger}$ = 51 kJ mol^{-1} is markedly lower than the AM1 barrier.

Overall it appears that the AM1/OPLS treatment of the solvated enzyme-substrate complex provides reasonably good geometries and predicts the important mechanistic features correctly, whereas calculated energies may contain errors. However, it is also evident that attempts to correct the deficiencies of the AM1 semiempirical method by means of a higher-level QM method (*e.g.* DFT) may not be successful unless they involve geometrical relaxation and consider not only E_{QM} but also E_{MM} and $E_{QM/MM}$; this is not feasible for free-energy calculations on systems of the size considered here.

6.2. Kinetic Isotope Effect Calculations

One of the most powerful experimental techniques for probing the nature of the TS is the measurement of a kinetic isotope effect (KIE), a rate constant ratio that arises from isotopic substitution at a particular position in a molecule (*17*). The magnitude and direction of a KIE contains information about the mechanistic events in a chemical reaction, reflecting differences in bonding between the reactants and the TS. In general, computational modelling of chemical mechanisms may bear little relation to experiment, but the use of KIEs provides a very strong anchor and a very stern test: a computed TS which predicts KIEs for multiple isotopic substitutions in agreement with reliable experimental results is likely to be of real value for rationalisation and prediction (*18*). Ten years ago, Ian Williams (Bath) and Steve Withers (University of British Columbia) agreed, respectively, to compute and to measure KIEs (on V_{max}/K_m) for hydrolysis of a range of isotopologues of p-nitrophenyl xylobioside catalysed by BCX (*19*); furthermore, they agreed that the experimental values would not be disclosed until after the computational modelling had been performed. The experiments were carried out some time ago, but it has taken until now to achieve the goal of calculating the KIEs for the enzyme-catalysed reaction. Although this work is still not complete, we present preliminary computational results for these KIEs to

provide more insight on the catalytic pathway and the transition states involved in the pathway for glycoside hydrolysis.

6.2.1. Computational Procedure

Traditionally, KIEs for a chemical reaction have been calculated using QM methods by computation of the optimised geometry and force constants for one reactant structure and one transition structure; moreover, solvation has been treated either by supermolecule calculations that ignored bulk solvent or by continuum methods that ignored specific interactions with the solute. In contrast, a recently developed methodology has been employed in this study that aims not only to treat explicitly solvation effects and/or interactions with the protein environment but also to take account of the multiple solvent and protein configurations accessible under normal experimental (*20*). For each of many "snapshots" from an MD trajectory for a particular temperature, we perform QM/MM geometry optimisation (to either a local minimum RC or saddle-point TS), compute a subset hessian for selected atoms and calculate the semi-classical KIE as the average of individual values obtained for all pairwise combinations of RCs and TSs.

The procedure adopted for this preliminary study of KIEs for the glycosylation step of p-nitrophenyl xylobioside hydrolysis catalysed BCX differs slightly from that outlined above, and is described in detail as follows.

1. The refined TS for the glycosylation step of the enzymatic reaction (see chapter 6, page 141-142, for more details) was followed for 100 ps of QM/MM molecular dynamics simulation at a temperature of 300 K. The ξ_1 and ξ_2 values were subjected to a very strong harmonic constraint (k_i = 20000 kJ mol^{-1} Å^{-2}) to keep the TS coordinate variables close to its reference value while the rest of the system was allowed to move freely.
2. For the purpose of this pilot study, a limited sample of just eight structures were selected along the overall dynamic run, each of which was refined to a first-order saddle point (a TS) using standard second-derivative saddle-point location techniques. The QM region together with 17 active-site residues of the immediately surrounding MM region (together the "core") were optimized

using second-derivative techniques, and the remainder of the system (the "environment") was optimized using a standard gradient minimizer. A hessian calculation was performed for every stationary point, for a subset of 386 mobile atoms corresponding to the core used in the saddle-point refinement.

3. The optimized structure of the substrate in water was taken as the appropriate reactant species from which to initiate an unconstrained 100 ps QM/MM MD trajectory. Again, eight structures were selected (as in step 2 above) and each was refined to a local energy minimum by means of a micro-iterative QM/MM geometry optimization for a core (comprising the substrate and a selection of MM water molecules with at least one atom within 5.5 Å of any atom of the substrate) and an environment (comprising the rest of the MM waters). The total number of waters in the core varied between 63 and 73 molecules. A hessian was determined for the core subset of mobile atoms (between 238 and 268) for each relaxed structure.
4. For each possible combination of a particular RC with a particular TS, an individual KIE was evaluated from the optimized geometries and computed hessians using the CAMVIB and CAMISO programmes (*21*, *22*). Each structure was considered as if it were a whole cluster with 3 translational, 3 rotational and 3N – 6 (or 3N – 7 for a TS) vibrational degrees of freedom. Small non-zero frequencies for translational and rotational motion were eliminated by a projection method implemented in CAMVIB, such that the resultant pure vibrational frequencies for isotopomeric species satisfied the Teller-Redlich product rule, being entirely consistent with the masses and moments of inertia obtained from the molecular geometries (*21*, *22*). Partition functions were evaluated at a temperature of 300 K within the harmonic-oscillator, rigid-rotor, ideal-gas approximations and were utilised within a standard semi-classical transition-state theoretical treatment of isotope effects (*17*).
5. A simple mean and standard deviation was determined from the $8 \times 8 = 64$ individual KIEs.

6.2.2. Results and Discussion

Figure 6.8 shows the QM/MM calculated KIEs on V_{max}/K_m for the following isotopic substitutions: 2 α-^{2}H, β-^{2}H, γ-$^{2}H_2$ and $^{18}O_{ring}$ and 1 α-^{13}C and $^{18}O_{lg}$. These values report on changes in bonding affecting the substituted positions in the substrate as between the free reactant state in aqueous solution and the transition state within the BCX active site.

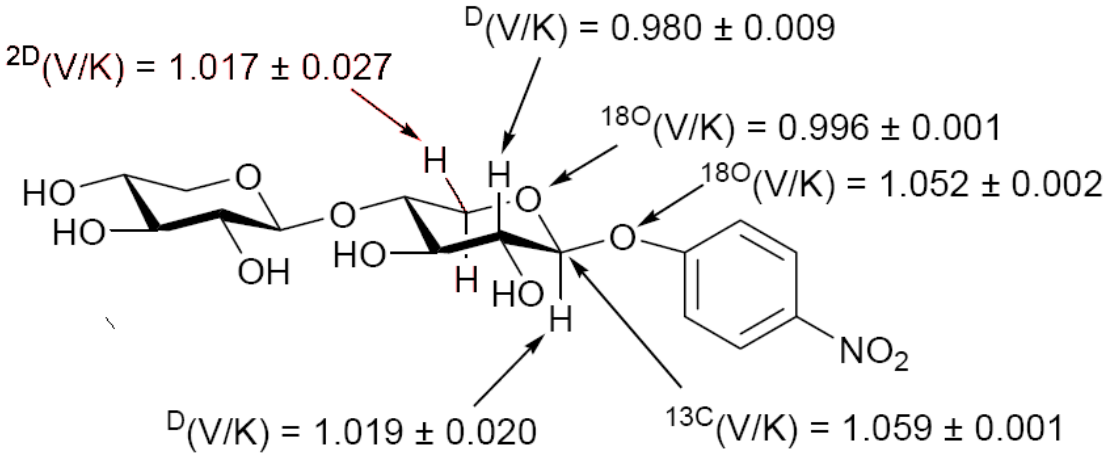


Fig. 6.8. Calculated kinetic isotope effects for hydrolysis of p-nitrophenyl xylobioside catalysed by wild-type BCX.

Each of the calculated KIEs has an entirely plausible value, and together they provide a picture of the transition state. At the time of writing we still do not know the experiment values for these KIEs, so it is not possible yet to comment upon the extent of agreement (or disagreement) between our computational model and experiment. However, we may note how the calculated KIE values would usually be interpreted if they were experimental values.

1 α-^{13}C: the large normal KIE (1.06) at the anomeric carbon is consistent with an S_N2-like transition state.

1 $^{18}O_{lg}$: similarly, the large normal KIE (1.05) at the leaving-group oxygen is consistent with the observation from the calculations that glycosidic bond is substantially broken in the transition state for glycosylation and that the p-nitrophenolate departs as the anion.

2 α-^{2}H: the rather small normal KIE (1.02) is also consistent with an S_N2-like transition state in which the "loosening" effect of glycosidic bond cleavage is largely offset by the "tightening" effect of covalent bond formation to the nucleophile.

2 β-^{2}H: the small inverse KIE (0.98) is somewhat unexpected but is consistent with the inductive effect of a single deuterium in the β-position adjacent to a developing carbocation centre (*23*). It suggests that there is no significant extent of hyperconjugation between the β-CH bond and the anomeric centre in the transition state: this could arise either because there is very little charge development at C_α or because the β-CH bond is approximately orthogonal to - and therefore unable to overlap significantly with – the incipient empty p-orbital.

2 $^{18}O_{ring}$: the very small inverse KIE (0.996) is consistent with a very small net change in bonding to this atom: even if there is conjugation between an electronic lone pair on O_{ring} and C_α, the "tightening effect of increased π-bonding could be offset by the change from sp^3 to sp^2 hybridisation.

2 γ-$^{2}H_2$: the small normal KIE (1.02 for two deuteriums) could indicate a small amount of hyperconjugation between these CH bonds and O_{ring}.

Of course, we already have direct knowledge of the transition structures determined for this reaction from the saddle-point searches, so it should be possible (once a more complete averaging over thermally accessible configurations has been performed) to assess the significance of the conventional interpretations of experimental KIEs. Furthermore, it may be possible to consider calculated KIEs for an alternative mechanism (e.g. proton transfer precedes nucleophilic substitution – as it would do for a poorer leaving group) and discriminate between these mechanisms on the basis of comparison with the actual experimental KIEs.

References

1. Field, M. J., Albe, M., Bret, C., Proust-De Martin, F., and Thomas, A. (2000) The Dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.,* **21**, 1088-1100.
2. Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. (1985) The development and use of quantum-mechanical molecular models.76. AM1 - a new general-purpose quantum-mechanical molecular model. *J. Am. Chem. Soc.,* **107**, 3902-3909.
3. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, **105**, 6474-6487.
4. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926-935.
5. Valleau, J. P., and Torrie, G. M. in *Statistical Mechanics*; Berne, B. J., Ed.; Plenum Press: NewYork, 1977; Part A, pp 169-194; Torrie, G. M., and Valleau, J. P. (1974) Monte-Carlo free-energy estimates using non-Boltzmann sampling - application to subcritical Lennard-Jones fluid, *Chem. Phys. Lett.*, **28**, 578-581.
6. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J. M. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.,* **13**, 1011-1021.
7. Roux, B. (1995) The calculation of the potential of mean force using computer-simulations. *Comput. Phys. Commun.,* **91**, 275-282.
8. Allen, M. P., and Tildesley, D. J. (1987) *Computer Simulation of Liquids*; Oxford University Press: Oxford, UK.

9. Schaftenaar, G., and Noordik, J. H. (2000) Molden: a pre- and post-processing program for molecular and electronic structures. *J. Comput.-Aided Mol. Design*., **14**, 123-134.

10. Gaussian 03, Revision C.02, Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., and Pople, J. A., Gaussian, Inc., Wallingford CT, 2004.

11. Becke, A. D. (1992) Density-functional thermochemistry.1. The effect of the exchange-only gradient correction. *J. Chem. Phys.*, **96**, 2155-2160.

12. Reed, A. E., Weinstock, R. B., and Weinhold, F. (1985) Natural population analysis. *J. Chem. Phys*., **83**, 735-746.

13. Lawson, S. L., Wakarchuk, W. W., and Withers, S. G. (1996) Effects of both shortening and lengthening the active site nucleophile of *Bacillus circulans* xylanase on catalytic activity. *Biochemistry,* **35**, 10110-10118.

14. Lawson, S. L., Wakarchuk, W. W., Withers, S. G. (1997) Positioning the acid/base catalyst in a glycosidase: Studies with *Bacillus circulans* xylanase. *Biochemistry*, **36**, 2257-2265.

15. Namchuk, M. N., and Withers, S. G. (1995) Mechanism of Agrobacterium β-glucosidase: kinetic analysis of the role of noncovalent enzyme/substrate interactions. *Biochemistry*, **34**, 16194-16202.

16. Wang, Q., Graham, R. W., Trimbur, D., Warren, R. A. J. , and Withers, S. G. (1994) Changing enzymatic reaction mechanisms: conversion of a retaining glucosidase to an inverting enzyme. *J. Am. Chem. Soc.*, **116**, 11594-11595.

17. *Isotope Effects in Chemical Reactions*, eds. Collins, C. J. and Bowman, N. S., Van Nostrand Reinhold, New York, 1970; Melander, L. and Saunders, W. H., *Reaction Rates of Isotopic Molecules*, Wiley, New York, 1980.

18. Barnes, J. A., and Williams, I. H. (1996), Theoretical modelling of kinetic isotope effects for glycoside hydrolysis in aqueous solution by a hybrid quantum-mechanical/molecular-mechanical method, *Chem. Commun.*, 193-194.

19. Withers, S. G. (1999) personal communication to Williams, I. H.

20. Kanaan, N., Ruiz-Pernía, J. J., and Williams, I. H. (2008) QM/MM simulations for methyl transfer in solution and catalysed by COMT: ensemble-averaging of kinetic isotope effects, *Chem. Commun.* 6114-6116.

21. Williams, I. H. (1982) On the representation of force fields for chemically reacting systems, *Chem. Phys. Lett.*, **88**, 462-466.

22. Williams, I. H. (1983) Force-constant computations in cartesian coordinates. Elimination of translational and rotational contributions, *J. Mol. Struct., THEOCHEM*, **94**, 275-284.

23. Williams, I. H. (1985) Theoretical studies of isotope effects pertinent to solvolysis mechanisms, *J. Chem. Soc., Chem. Commun.* 510-511.

7. Role of Tyr69

7.1. Introduction

In the non-covalent complex, Tyr69 donates a strong hydrogen bond to the nucleophilic oxygen atom ($O^{\epsilon 2}$) of Glu78 (hereinafter denoted O_{nuc}; see Fig. 6.3) and accepts a hydrogen bond from the hydroxyl group at the 2-position of the distal xylose moiety. In contrast, in the covalent intermediate, the hydrogen bond donated to the O_{nuc} is weaker, consistent with the ether character of its partner, and a stronger interaction is formed between Tyr69 and the ring oxygen (O_{ring}) of the proximal xylose moiety. The phenolic oxygen (O_Y) of Tyr69 is very important for catalysis, as evidenced by the observation that the Y69F mutant exhibits no detectable enzyme activity (*1*). The nature of this $O_YH_Y\cdots O_{ring}$ interaction is therefore an intriguing question that we seek to investigate here. In chapter 5 we shed some light upon the role of Tyr69 in stabilizing the $^{2,5}B$ conformation of the distorted proximal sugar ring of a xylobioside substrate, but in this chapter we have employed computational modeling using quantum mechanical/molecular mechanical (QM/MM) potentials to simulate free-energy changes occurring in the glycosylation step of the mechanism. We aim to obtain deeper understanding of the catalytic role of Tyr69 and we achieve this by comparative analysis of the transition states involved in the reaction pathway for the wild-type BCX enzyme and the Y69F mutant.

7.2. Computational procedure

Model system preparation, QM/MM portioning, AM1/OPLS MD trajectories, optimizations of energy minima and first-order saddle points, and PMFs were computed as described in chapter 6. Atomic charges were obtained by natural population analysis (*2*), as implemented in Gaussian03; single-point calculations were performed using an enlarged QM region that included Tyr69 (or Phe in the Y69F mutant) and averages were taken over ten structures randomly selected from the AM1/OPLS MD trajectories.

7.3. Results and Discussion

In order to obtain insight into the precise role of Tyr69 upon the catalysis, it is instructive to set up a comparison between the wild-type enzyme and Y69F mutant. Our comparative study is mainly based upon three main axes - hydrogen bond analysis, free-energy profiles for wild-type and Y69F mutant and transition state analysis for wild-type and Y69F mutant.

Hydrogen bond analysis

We showed previously in chapter 4 that the phenolic OH group of Tyr69 donates a hydrogen bond to either O_{nuc} of Glu78 or O_{ring} of the proximal xylose, fluctuating between the two during the course of an AM1/OPLS MD trajectory for the non-covalent RC of wild-type BCX. PMF calculations revealed that these interactions stabilize the $^{2,5}B$ boat conformation of the sugar ring by about 20 kJ mol^{-1} relative to the $^{4}C_1$ chair conformation, thereby favouring an approximately coplanar alignment of the atoms about the C_{ano}–O_{ring} bond in the proximal ring of the substrate, as required for the oxacarbenium-like TS.

Figure 7.1 extends this same analysis to the TS (TS1) and product (INT) of the nucleophilic substitution step of glycosylation. During the course of a 30 ps AM1/OPLS MD trajectory for the TS, it is apparent that the $H_Y\cdots O_{nuc}$ distance is consistently shorter than the $H_Y\cdots O_{ring}$ distance, indicating that the hydrogen bond between Tyr69 and Glu78 is favoured, although both distances are shorter than the averages found in RC. However, in the covalently-bonded INT the $H_Y\cdots O_{nuc}$ distance is consistently longer than the $H_Y\cdots O_{ring}$ distance, indicating that Tyr69 now donates its hydrogen bond exclusively to the xylose ring rather than to Glu78, although the average distance to the latter is similar to that in RC.

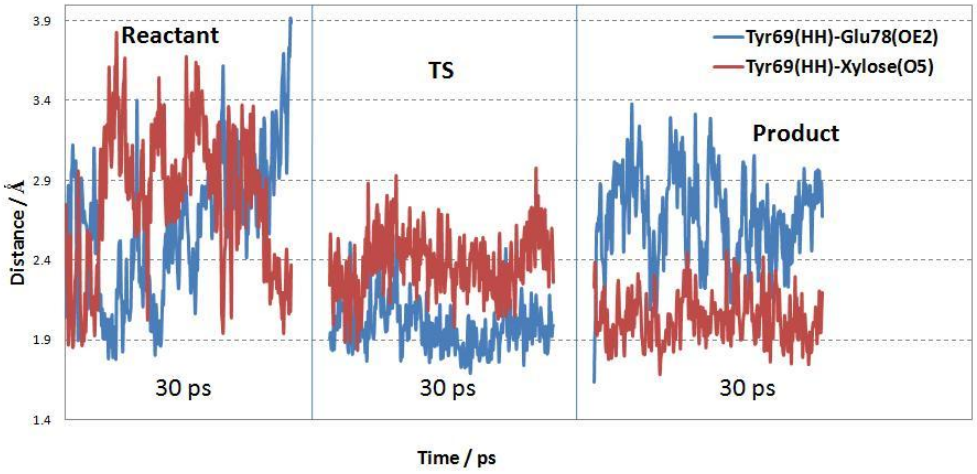


Figure 7.1. Plot of the hydrogen bonding distances $H_Y \cdots O_{nuc}$ (blue) between Tyr69 and Glu78 and $H_Y \cdots O_{ring}$ (red) between Tyr69 and xylose during AM1/OPLS MD trajectories for reactant, TS and product of the nucleophilic substitution step for wild-type BCX

Figure 7.2 shows the $O_Y \cdots O_{ring}$ distance over 30 ps AM1/OPLS MD trajectories for the RC, TS (TS1) and product (INT) of nucleophilic substitution. It is evident that the average distance in both the TS and the covalently-bonded INT is similar and significantly shorter than in the non-covalent reactant complex. The catalytic implications of this close approach between Tyr69 and the xylose ring in the TS are considered below.

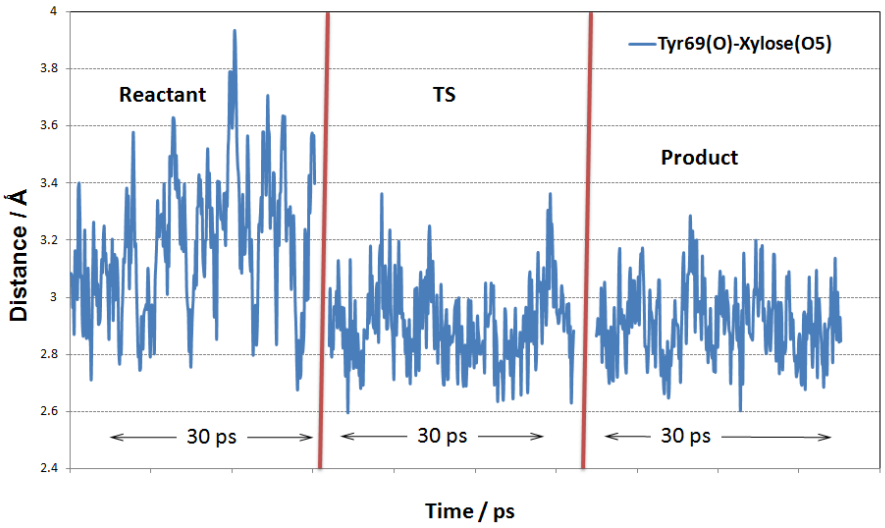


Figure 7.2. A plot of the $H_Y \cdots O_{ring}$ distance between Tyr69 and the proximal xylose residue during AM1/OPLS MD trajectories for reactant, TS and product of the nucleophilic substitution step for wild-type BCX:

Free-energy profiles for wild-type and Y69F mutant

To further understand the catalytic function of Tyr69, it is instructive to compare AM1/OPLS PMFs calculated for reactions of both wild-type BCX and the Y69F mutant. Since the free-energy profiles for the proton transfer step did not show any noticeable difference, only those for the nucleophilic substitution step are shown in Figure 8. The shapes of the profiles are similar, but the free-energy barrier for reaction of the Y69F mutant is higher by ~ 42 kJ mol^{-1}; this corresponds to a factor of ~2×10^7 in k_{cat} at 25 °C, in accord with the observation that the Y69F mutant has less than 0.01% of the activity of wild-type BCX (*1*). Note that the two systems differ only by a single atom, O_Y. Although the absolute magnitudes of each barrier are undoubtedly overestimated by the AM1 description of the QM region, most of the systematic errors should cancel in the comparison. The source of the difference must lie in the interaction between Tyr69 and the substrate. We anticipate that the change from tyrosine to phenylalanine (both in the MM region) will polarize higher-level QM wavefunctions in a qualitatively similar (though undoubtedly quantitatively different) fashion to AM1. We therefore consider our AM1/OPLS estimate of the difference in the free-energy barrier for glycosylation to be at least qualitatively reliable. A quantitative assessment of error, as compared to an appropriate high-level QM method, would be meaningless unless it were evaluated for completely converged populations of conformations accessible under condensed-phase conditions, since there are large fluctuations between energy differences taken at arbitrary "snapshot" structures taken from the MD trajectories.

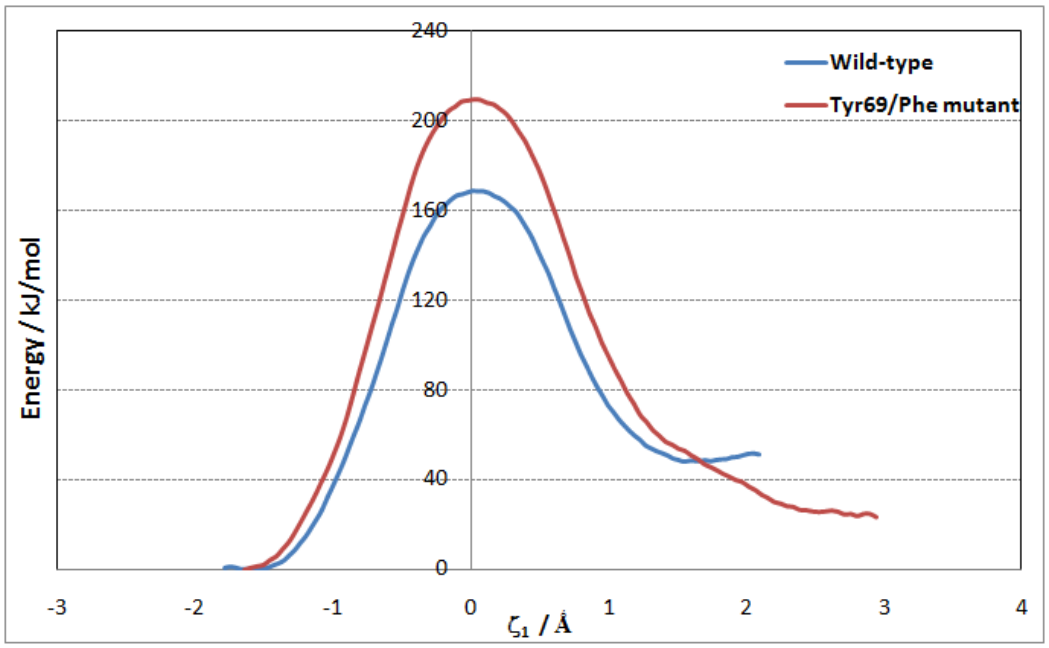


Figure 7.3. Free-energy profiles for formation of the glycosyl-enzyme intermediate in wild-type BCX (black) and Y69F mutant (blue).

Transition state analysis for wild-type and Y69F mutant

In order to obtain insight into the changes occurring in the electronic charge distribution as between the RC and the TS1 in the wild-type and mutant enzymes, detailed analysis was performed on sets of ten "snapshot" structures selected randomly from along the 30 ps QM/MM MD trajectories. Average interatomic distances (Table 7.1) and atomic charges (Table 7.2) were computed with the B3LYP/6-31+G* method for a QM region which was extended to include the active-site tyrosine or phenylalanine. It is important to note that the mutation causes no significant geometrical perturbation in the active site region (Figure 7.4): the position of Glu78 remains unchanged, implying that Tyr69 plays no role in placing the Glu78 in a position appropriate for nucleophilic attack, in agreement with earlier X-ray crystallographic analysis (*3*).

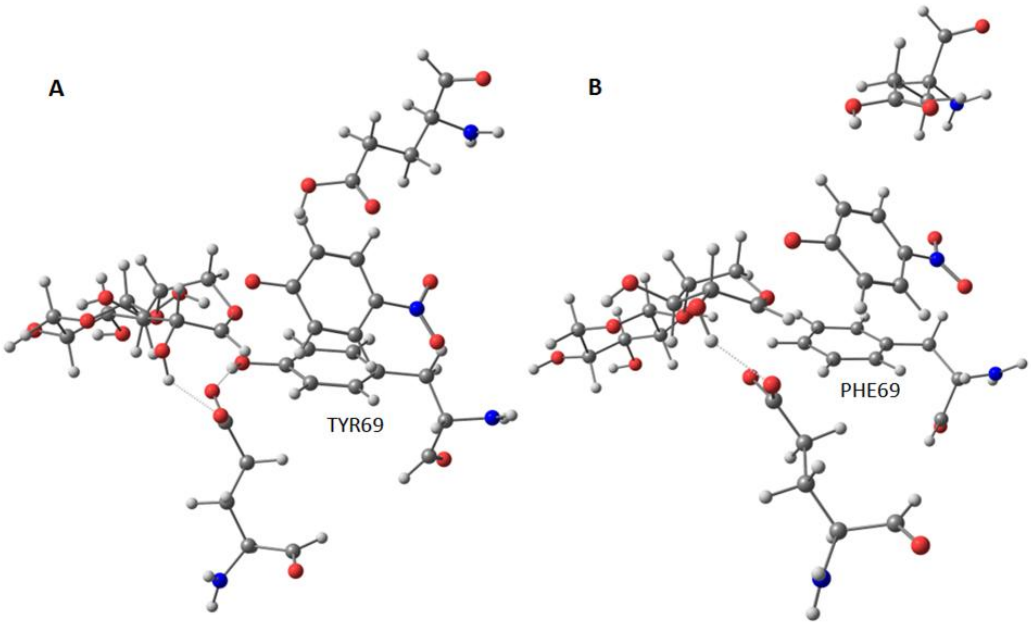


Figure 7.4. Representative B3LYP/6-31+G* optimized transition structures selected for atomic charge calculations for wild-type BCX (A) and the Y69F mutant (B), showing the full extent of the QM region

Table 7.1. Interatomic distances (Å) and transition frequencies for first-order saddle points optimized AM1/OPLS the transition states of wild-type BCX and the Y69F mutant.

	Wild-type	Y69F
$C_{ano}\cdots O_{lg}$	2.127	2.150
$O_{nuc}\cdots C_{ano}$	2.110	2.129
$H_Y\cdots O_{nuc}$	1.909	
$O_Y\cdots O_{ring}$	2.829	
$\nu^{\ddagger}/cm^{-1}$	460*i*	422*i*

Table 7.2. Average B3LYP/6-31+G* atomic charges (|*e*|) on selected atoms of first-order saddle points optimized from structures selected randomly along AM1/OPLS MD trajectories for the transition states of wild-type BCX and the Y69F mutant.

	AM1/OPLS				B3LYP/6-31+G*			
	Wild-type		Y69F		Wild-type		Y69F	
	R	TS	R	TS	R	TS	R	TS
H_{ano}	0.196	0.246	0.210	0.258	0.138	0.190	0.152	0.190
C_{ano}	0.111	0.318	0.117	0.312	0.350	0.353	0.339	0.347
O_{ring}	–0.273	–0.128	–0.258	–0.120	–0.489	–0.431	–0.502	–0.411
O_{lg}	–0.272	–0.593	–0.263	–0.580	–0.551	–0.672	–0.470	–0.658
O_Y	–0.585	–0.619			–0.604	–0.619		
H_Y	0.435	0.366			0.350	0.366		

Zechel and Withers suggested that in the TS there might be a ‘stabilizing electrostatic or dipolar interaction’ between O_Y and the partial positive charge on O_{ring}, which they depicted graphically as an arrow from an electronic lone pair on the former towards the latter (*4*). Their description may now be re-interpreted in the light of the atomic charges now computed for the RC and the TS, which have been averaged over a selection of structures from MD trajectories for these species. Table 3 contains average charges obtained from natural population analysis of the calculated electron density distributions. Regardless of which method is used to define atomic charges (alternatives include Mulliken and electrostatic potential charges), oxygen atoms invariably carry partial negative charges. Thus, as the glycosidic bond undergoes cleavage, and the glycon acquires oxacarbenium-ion character, so the charge on O_{ring} becomes less negative rather than actually positive. Considering C_{ano} and H_{ano} as a unit, the charge separation along the O_{ring}–$(CH)_{ano}$ bond differs little between the RC (0.977 |e|) and the TS (0.974 |e|) for wild-type BCX, but is matched by a

complementary separation of charge along the O_Y–H_Y bond as between the RC (0.954) and TS (0.985), corresponding to a pair of roughly antiparallel bond dipoles. In the Y69F there is no O_Y–H_Y bond dipole present to stabilize the O_{ring}–$(CH)_{ano}$ bond dipole.

References

1. Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J., and Yaguchi, M. (1994) Mutational and crystallographic analyses of the Active-site residues of the *Bacillus circulans* xylanase. *Protein Sci.*, **3**, 467-475.
2. Reed, A. E., Weinstock, R. B., and Weinhold, F. (1985) Natural population analysis. *J. Chem. Phys*., **83**, 735-746.
3. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase. *Biochemistry*, **38**, 5346-5354.
4. Zechel, D. L., and Withers, S. G. (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.*, **33**, 11-18.

8. Ongoing and Future Work

Overview

In this chapter, we not only present some of the calculations that still are ongoing and the results we have obtained so far, but also we report some other future plans that might help provide more insight into the structural and mechanistic features for BCX.

8.1. Ongoing work

8.1.1. Positioning of the catalytic residues in BCX

Chapter 3 provides detailed information about kinetic studies performed to probe the influence of both shortening and lengthening of the active site catalytic residues of endo-β-1,4-xylanase from *Bacillus circulans*. Positional changes of acid/base residue, Glu172, did not have a remarkable effect on the catalytic activity. However, shortening of the nucleophilic side chain in a Glu78Asp mutant decreased k_{cat}/K_m values at least 1600 fold for the aryl xylobiosides and much more for xylan, whereas extending the side chain in a carboxymethylated-Glu78Cys mutant decreased k_{cat}/K_m values by 16 – 100 fold (*1-3*). In other words, the positional requirements for proton transfer are less demanding than those for nucleophilic attack.

In this chapter, and in light of aforementioned findings, we perform mutational studies for the nucleophilic residue, Glu78, aiming to provide more reasonable rationalization for such observations.

However, we are still in early stages to provide a full picture in that regard. Below are the computational procedure and results we have obtained at this stage so far.

8.1.1.1. Computational procedure

Glu78/Asp mutant model preparation

Mutations have been done manually by altering the Glu78 coordinates into Asp residue. Retaining the original orientation of the Glu78 was carefully monitored

during the modification of Glu78 to Asp78. In contrast to software-induced mutation, manual mutation has been found to be a good tactic to keep the orientation and the integrity of the active site architecture. System solvation, optimization and equilibration have been achieved as described in chapter 4.

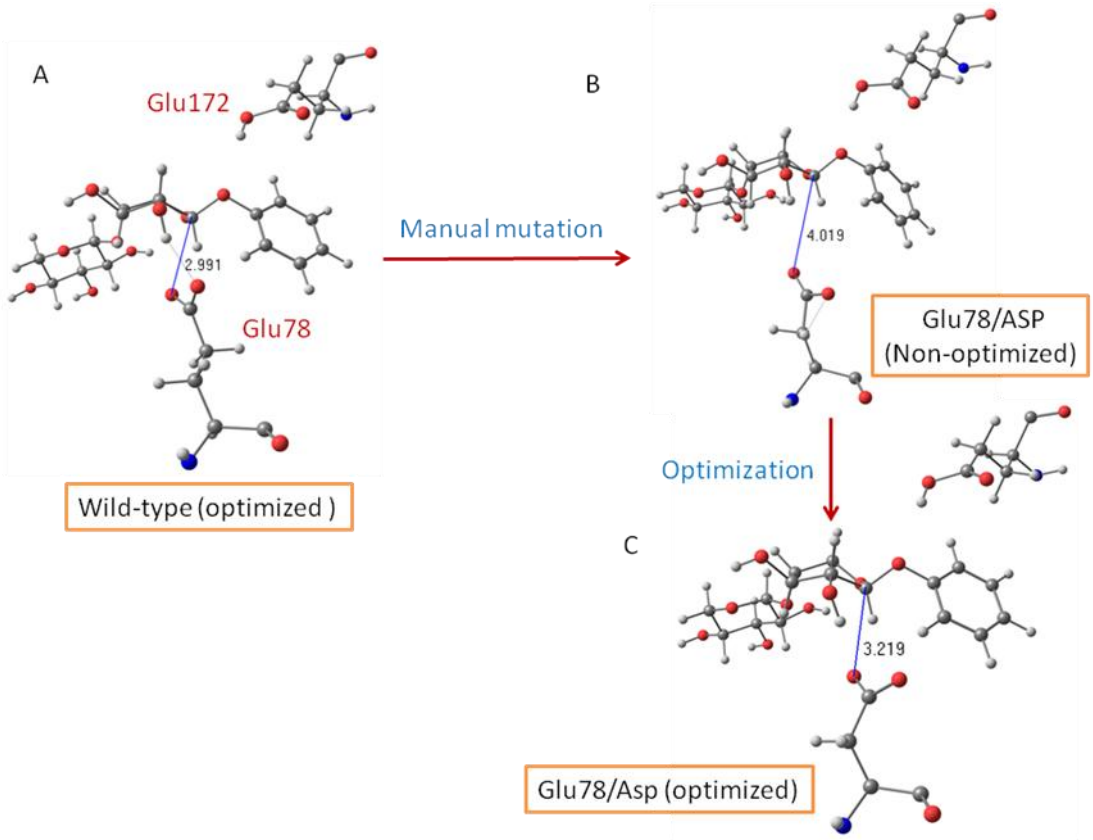


Fig.8.1. Glu78Asp mutant model preparation and optimization.

Free-energy calculations

The free-energy profile is computed as explained previously in chapter 6.

8.1.1.2. Results and conclusions

The active site after optimization

Despite chain shortening of the nucleophilic residue (Glu78 →Asp), the distance between the anomeric carbon of the proximal xylose ring and the nucleophilic oxygen of Asp78 still maintains nearly the same value as measured in the wild-type enzyme (3.40 Å and 3.21 Å for the wild-type and Glu78Asp mutant, respectively).

The noticeable decrease in the distance between the anomeric carbon of the proximal xylose ring and the nucleophilic oxygen of Asp78 after optimization, 3.21 Å compared to 4.01 Å before optimization, could be a result of either the movement of the substrate towards the nucleophilic oxygen of Asp78 or a change in orientation of active site residues.

To figure out why the distance between the anomeric carbon of the proximal xylose ring and the nucleophilic oxygen of Asp78 remains unaffected after chain shortening of the nucleophilic residue, we are currently computing the root mean square deviation, RMSD, for the optimized system coordinates relative to the starting configuration. More computations in the same regard are still going on.

The overall distance between the two catalytic residues in the wild-type enzyme and Glu78/Asp is 7.02 Å and 9.32 Å, respectively.

Free-energy profile for the wild-type and Glu78Asp mutant

Figure 8.2 shows that the computed free-energy profiles for the wild-type and Glu78/Asp are in a great accord with the experimental observations. Glu78Asp mutation increases the free-energy barrier for the glycosylation reaction by 40 kJ mol–1.

Further calculations, to provide more understanding, are ongoing.

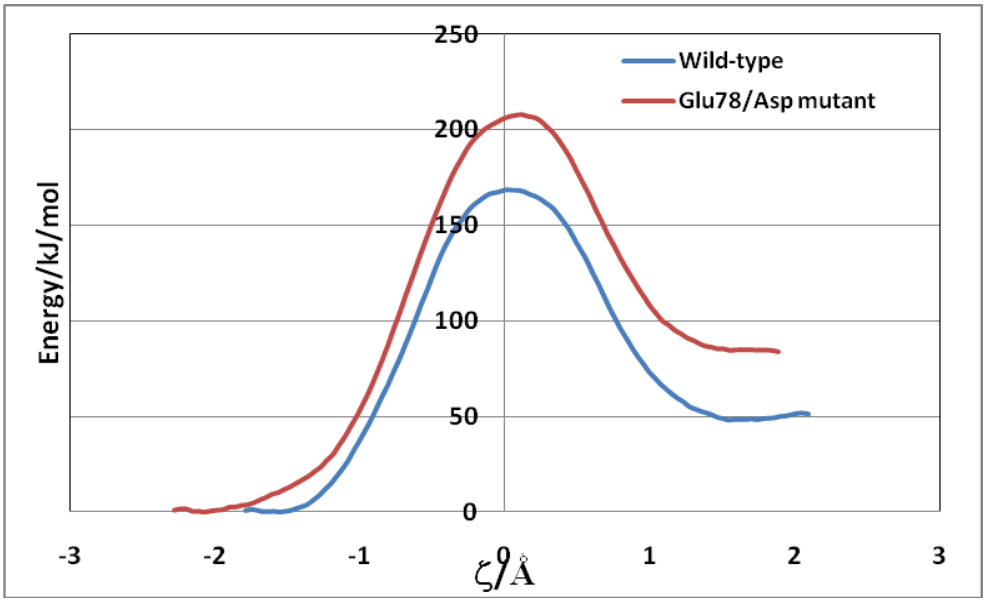


Figure 8.2. Free-energy profiles for formation of the glycosyl-enzyme intermediate in wild-type BCX (blue) and Glu78/Asp mutant (red).

8.2. Future work

However, the work achieved in this thesis provides valuable information for many structural and mechanistic features of BCX, there are still many points of interest in BCX that need more investigation.

In this thesis, and to understand the catalytic mechanism, PNP as a leaving group has been used to understand the catalytic mechanism to validate our model with respect to the experimental and kinetic studies. With a great deal of evidence, our model system is considered as a good complement to the experiments. In the next stage, simulation of the natural leaving group, xylose, could be instructive to help understand the genuine catalytic process and to see the effect of pK_a variation on the energy landscape.

In this thesis, we have been focusing exclusively on the glycosylation process. To obtain a complete frame of the catalytic pathway, exploration of the deglycosylation process would be necessary.

The number of water molecules involved in the catalytic process has been a subject of argument (*4*). Simulation of a model system with different number of water molecules in the active site will be insightful to fully understand how critical is the number and configuration of the water molecules in the active site.

The role of Tyr80 in the deglycosylation step is critical as evident by the fact that the mutation of Tyr80 leaves the enzyme with less than 0.045% of its wild-type activity (*5*). Tyr80 plays a crucial role in the orientation of the water molecules in the active site (*4*). Further QM/MM simulations on the role of Tyr80 on catalysis are still needed.

References

1. Davies, G., and Henrissat, B. (1995) Structures and Mechanisms of Glycosyl Hydrolases, *Structure*, **3**, 853-859.
2. Lawson, S. L., Wakarchuk, W. W., and Withers, S. G. (1997) Positioning the acid/base catalyst in a glycosidase: studies with *Bacillus circulans* xylanase, *Biochemistry*, **36**, 2257-2265.
3. Lawson, S. L., Wakarchuk, W. W., and Withers, S. G. (1996) Effects of both shortening and lengthening the active site nucleophile of *Bacillus circulans* xylanase on catalytic activity, *Biochemistry*, **35**, 10110-10118.
4. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase, *Biochemistry*, **38**, 5346-5354.
5. Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J., and Yaguchi, M. (1994) Mutational and Crystallographic Analyses of the Active-Site Residues of the Bacillus Circulans Xylanase, *Protein Sci.*, **3**, 467-475.